

IJCSIS Vol. 15 No. 12 (I), December 2017
ISSN 1947-5500

**International Journal of
Computer Science
& Information Security**

© IJCSIS PUBLICATION 2017
Pennsylvania, USA

Indexed and technically co-sponsored by :



AUTHOR SERIES



IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

CALL FOR PAPERS

International Journal of Computer Science and Information Security (IJCSIS) January-December 2017 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.

See authors guide for manuscript preparation and submission guidelines.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

Deadline: see web site

Notification: see web site

Revision: see web site

Publication: see web site

Context-aware systems
Networking technologies
Security in network, systems, and applications
Evolutionary computation
Industrial systems
Evolutionary computation
Autonomic and autonomous systems
Bio-technologies
Knowledge data systems
Mobile and distance education
Intelligent techniques, logics and systems
Knowledge processing
Information technologies
Internet and web technologies
Digital information processing
Cognitive science and knowledge

Agent-based systems
Mobility and multimedia systems
Systems performance
Networking and telecommunications
Software development and deployment
Knowledge virtualization
Systems and networks on the chip
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

For more topics, please see web site <https://sites.google.com/site/ijcsis/>

arXiv.org

Google scholar

SCIRUS
search engine for science

ScientificCommons

Scribd

.docstoc
find and share professional documents

BASE
Bielefeld Academic Search Engine

CiteSeer^x beta

dblp.uni-trier.de
Computer Science
Bibliography

DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS



ProQuest

For more information, please visit the journal website (<https://sites.google.com/site/ijcsis/>)

Editorial Message from Editorial Board

It is our great pleasure to present the **December 2017 issue** (Volume 15 Number 12) of the **International Journal of Computer Science and Information Security (IJCSIS)**. High quality research, survey & review articles are proposed from experts in the field, promoting insight and understanding of the state of the art, and trends in computer science and technology. It especially provides a platform for high-caliber academics, practitioners and PhD/Doctoral graduates to publish completed work and latest research outcomes. According to Google Scholar, up to now papers published in IJCSIS have been cited over 9800 times and this journal is experiencing steady and healthy growth. Google statistics shows that IJCSIS has established the first step to be an international and prestigious journal in the field of Computer Science and Information Security. There have been many improvements to the processing of papers; we have also witnessed a significant growth in interest through a higher number of submissions as well as through the breadth and quality of those submissions. IJCSIS is indexed in major academic/scientific databases and important repositories, such as: Google Scholar, Thomson Reuters, ArXiv, CiteSeerX, Cornell's University Library, Ei Compendex, ISI Scopus, DBLP, DOAJ, ProQuest, ResearchGate, Academia.edu and EBSCO among others.

A great journal cannot be made great without a dedicated editorial team of editors and reviewers. On behalf of IJCSIS community and the sponsors, we congratulate the authors and thank the reviewers for their outstanding efforts to review and recommend high quality papers for publication. In particular, we would like to thank the international academia and researchers for continued support by citing papers published in IJCSIS. Without their sustained and unselfish commitments, IJCSIS would not have achieved its current premier status, making sure we deliver high-quality content to our readers in a timely fashion.

"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication." We would like to thank you, the authors and readers, the content providers and consumers, who have made this journal the best possible.

For further questions or other suggestions please do not hesitate to contact us at ijcsiseditor@gmail.com.

A complete list of journals can be found at:
<http://sites.google.com/site/ijcsis/>

IJCSIS Vol. 15, No. 12, December 2017 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):



Open Access This Journal is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source.



Bibliographic Information

ISSN: 1947-5500

Monthly publication (Regular Special Issues)
Commenced Publication since May 2009

Editorial / Paper Submissions:

IJCSIS Managing Editor

ijcsiseditor@gmail.com

Pennsylvania, USA

Tel: +1 412 390 5159

IJCSIS EDITORIAL BOARD

IJCSIS Editorial Board	IJCSIS Guest Editors / Associate Editors
Dr. Shimon K. Modi [Profile] Director of Research BSPA Labs, Purdue University, USA	Dr Riktesh Srivastava [Profile] Associate Professor, Information Systems, Skyline University College, Sharjah, PO 1797, UAE
Professor Ying Yang, PhD. [Profile] Computer Science Department, Yale University, USA	Dr. Jianguo Ding [Profile] Norwegian University of Science and Technology (NTNU), Norway
Professor Hamid Reza Naji, PhD. [Profile] Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran	Dr. Naseer Alquraishi [Profile] University of Wasit, Iraq
Professor Yong Li, PhD. [Profile] School of Electronic and Information Engineering, Beijing Jiaotong University, P. R. China	Dr. Kai Cong [Profile] Intel Corporation, & Computer Science Department, Portland State University, USA
Professor Mokhtar Beldjehem, PhD. [Profile] Sainte-Anne University, Halifax, NS, Canada	Dr. Omar A. Alzubi [Profile] Al-Balqa Applied University (BAU), Jordan
Professor Yousef Farhaoui, PhD. Department of Computer Science, Moulay Ismail University, Morocco	Dr. Jorge A. Ruiz-Vanoye [Profile] Universidad Autónoma del Estado de Morelos, Mexico
Dr. Alex Pappachen James [Profile] Queensland Micro-nanotechnology center, Griffith University, Australia	Prof. Ning Xu, Wuhan University of Technology, China
Professor Sanjay Jasola [Profile] Gautam Buddha University	Dr . Bilal Alatas [Profile] Department of Software Engineering, Firat University, Turkey
Dr. Siddhivinayak Kulkarni [Profile] University of Ballarat, Ballarat, Victoria, Australia	Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece
Dr. Reza Ebrahimi Atani [Profile] University of Guilan, Iran	Dr Venu Kuthadi [Profile] University of Johannesburg, Johannesburg, RSA
Dr. Dong Zhang [Profile] University of Central Florida, USA	Dr. Zhihan Iv [Profile] Chinese Academy of Science, China
Dr. Vahid Esmaeelzadeh [Profile] Iran University of Science and Technology	Prof. Ghulam Qasim [Profile] University of Engineering and Technology, Peshawar, Pakistan
Dr. Jiliang Zhang [Profile] Northeastern University, China	Prof. Dr. Maqbool Uddin Shaikh [Profile] Preston University, Islamabad, Pakistan
Dr. Jacek M. Czerniak [Profile] Casimir the Great University in Bydgoszcz, Poland	Dr. Musa Peker [Profile] Faculty of Technology, Mugla Sitki Kocman University, Turkey
Dr. Binh P. Nguyen [Profile] National University of Singapore	Dr. Wencan Luo [Profile] University of Pittsburgh, US
Professor Seifeidne Kadry [Profile] American University of the Middle East, Kuwait	Dr. Ijaz Ali Shoukat [Profile] King Saud University, Saudi Arabia
Dr. Riccardo Colella [Profile] University of Salento, Italy	Dr. Yilun Shang [Profile] Tongji University, Shanghai, China
Dr. Sedat Akleyek [Profile] Ondokuz Mayıs University, Turkey	Dr. Sachin Kumar [Profile] Indian Institute of Technology (IIT) Roorkee

Dr Basit Shahzad [Profile] King Saud University, Riyadh - Saudi Arabia	
Dr. Sherzod Turaev [Profile] International Islamic University Malaysia	

TABLE OF CONTENTS

1. PaperID 30111701: Risk Management Process Analysis for Information & Communication Technology (ICT) Systems: Risk Assessment Perspective (pp. 1-6)

*(1) Nwagu, Chikezie Kenneth; (2) Omankwu, Obinnaya Chinecherem; (3) Okonkwo, Obikwelu Raphael
(1) Computer Science Department, Nnamdi Azikiwe University Awka Anambra State, Nigeria.
(2) Computer Science Department, Michael Okpara University of Agriculture Umudike Umuahia, Abia State, Nigeria.
(3) Computer Science Department, Nnamdi Azikiwe University Awka Anambra State, Nigeria.*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

2. PaperID 30111702: Evolutionary Perspective of Robotics (pp. 7-12)

*(1) Nwagu, Chikezie Kenneth; (2) Omankwu, Obinnaya Chinecherem; (3) Inyiama, Hycient
(1) Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,
(2) Computer Science Department, Michael Okpara University of Agriculture, Umudike Umuahia, Abia State, Nigeria
(3) Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria.*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

3. PaperID 30111703: Knowledge Discovery in Databases (KDD): An Overview (pp. 13-16)

*(1) Nwagu, Chikezie Kenneth; (2) Omankwu, Obinnaya Chinecherem; (3) Inyiama, Hycient
(1) Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,
(2) Computer Science Department, Michael Okpara University of Agriculture, Umudike Umuahia, Abia State, Nigeria
(3) Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria.*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

4. PaperID 30111704: Knowledge Management (KM): An Overview (pp. 17-21)

*(1) Omankwu, Obinnaya Chinecherem; (2) Nwagu, Chikezie Kenneth; (3) Inyiama, Hycient
(1) Computer Science Department, Michael Okpara University of Agriculture, Umudike Umuahia, Abia State, Nigeria
(2) Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,
(3) Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria.*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

5. PaperID 30111705: Perspectives for the Use of KM in Health Care (pp. 22-25)

*(1) Omankwu, Obinnaya Chinecherem; (2) Nwagu, Chikezie Kenneth; (3) Inyiama, Hycient
(1) Computer Science Department, Michael Okpara University of Agriculture, Umudike Umuahia, Abia State, Nigeria
(2) Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,
(3) Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria.*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

6. PaperID 30111706: Robotics and Artificial Intelligence: Differences and Similarities (pp. 26-28)

*(1) Omankwu, Obinnaya Chinecherem; (2) Nwagu, Chikezie Kenneth; (3) Inyama, Hycient
(1) Computer Science Department, Michael Okpara University of Agriculture, Umudike Umuahia, Abia State, Nigeria
(2) Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,
(3) Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria.*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

7. PaperID 30111707: Faculty Assignment and Timetabling using Optimisation (pp. 29-35)

*MA Nang Laik, Senior Lecturer, School of Business, Singapore University of Social Sciences, Singapore
SEN Prabir, Director Research, STATGRAF Research, Vancouver, Canada*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

8. PaperID 30111709: A Novel and Robust Retroareolar Region Detection in Mammography Images (pp. 36-39)

*Dr. P. Pedda Sadhu Naik, Professor & HOD, Dr. Samuel George Institute of Engineering & Technology, Markapur, AP.
Dr T. Venugopal, Professor & HOD, JNTUH College of Engineering, Jagityala, Telangana*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

9. PaperID 30111712: Improving DBMS Security through the use of a Checksum Technique (pp. 40-48)

*Naim Ajlouni *, Firas Ajlouni **, Alaa Ali Hameed ***
Istanbul Aydin University * ***
Lancashire College of Further Education ***

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

10. PaperID 30111713: Coupling GIS-based MCA and AHP techniques for Hospital Site Selection (pp. 49-56)

*Aseel B. Kmail, Jacquleen Jubran, Walid Sabbah
Geographic Information System Department, AAUJ, Jenin, Palestine*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

11. PaperID 30111714: Big and Connected Data Analysis with Graph and Relational Databases Using Collaborative Filtering Technique (pp. 57-65)

*Joseph George Davis, Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
Mr Joseph K. Panford, Senior Lecturer, Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana*

Dr James B. Hayfron-Acquah, Senior Lecturer, Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

12. PaperID 30111715: 3D Facial Features in Neuro Fuzzy Model for Predictive Grading Of Childhood Autism (pp. 66-75)

*Reji R, Research Scholar, School of Computer Sciences, Mahatma Gandhi University, Kottayam, Kerala, India.
Dr. P. SojanLal, Principal, Mar-Baselious Institute of Technology & Science, Kothamangalam, Kerala, India.*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

13. PaperID 30111718: Classification of Matrix Multiplication Methods Used to Encrypt-Decrypt Color Image (pp. 76-85)

Saleh A. Khawatreh, Computer Engineering Dept. Faculty of Engineering, Al-Ahliyya Amman University

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

14. PaperID 30111722: Document Processing and Arabic Optical Character Recognition: A User Perspective (pp. 86-97)

Yasser M. Alginahi, Senior Member, IEEE

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

15. PaperID 30111723: Digital Holy Quran Authentication, Certification and Standardization: A Survey (pp. 98-117)

*Yasser M. Alginahi (1,2), Muhammad Khurram Khan (3), Omar Tayan (2, 4), Mohammed Zakariah (5),
(1) Dept. of Computer Science, Deanship of Academic Services, Taibah University, KSA
(2) IT Research Center for the Holy Quran and Its Sciences (NOOR), Taibah University, KSA
(3) Center of Excellence in Information Assurance (CoEIA), King Saud University, KSA
(4) College of Computer Science and Engineering, Dept. of Computer Engineering, Taibah University, KSA
(5) Research Center of College of Computer and Information Sciences, King Saud University, KSA*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

16. PaperID 30111727: Proposed Architecture for Improving Security and Consistency of Data Transactions in Cloud Database using Tree-Based Consistency Approach (pp. 118-126)

*J. Antony John Prabu, Research Scholar and Assistant Professor, Department of Computer Science, St. Joseph's College, Trichy, Tamilnadu, India
Dr. S. Britto Ramesh Kumar, Assistant Professor, Department of Computer Science, St. Joseph's College, Trichy, Tamilnadu, India*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

17. PaperID 30111729: Color Histogram with Curvelet and CEDD for Content-Based Image Retrieval (pp. 127-143)

Heba A. Elnemr, Computers and Systems Department, Electronics Research Institute, Elbehoos St., Dokki, Giza, Egypt

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

18. PaperID 30111732: A Survey on Multi-Objective Task Scheduling Algorithm in Cloud Environment (pp. 144-151)

*Tanvi Gupta, Manav Rachna International Institute of Research and Studies
Dr. S. S. Handa, SRM University
Dr. Supriya Panda, Manav Rachna International Institute of Research and Studies*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

19. PaperID 30111736: Novel Evaluation of Protein Convertase Subtilisin/Kexin Type 9 (PCSK9) Gene by Motif Detection Technique (pp. 152-154)

*Raiha Tallat, CS. Department, (CIIT) Sahiwal, Pakistan
M. Shoaib, CS. Department, (CIIT) Sahiwal, Pakistan
Javed Ferzund, CS. Department, (CIIT) Sahiwal, Pakistan
Ahmad Nawaz Zaheer, MS. Department, (GCUF) Sahiwal, Pakistan
Sana Yaseen, CS. Department, (CIIT) Sahiwal, Sahiwal, Pakistan
Umar Draz, CS. Department, (CIIT) Sahiwal, Sahiwal, Pakistan*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

20. PaperID 30111737: A Review on Detection and Counter Measures of Wormhole Attack in Wireless Sensor Network (pp. 155-162)

Rabia Arshad & Saba Zia

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

21. PaperID 30111739: Performance Analysis of Spot diffusion Technique for Indoor Optical Communication System Employing LDPC (pp. 163-169)

*Sabira Khanam Shorna, Department of Computer Science & Engineering, Primeasia University, Dhaka, Bangladesh
Koushik Chandra Howlader, Dept. of CSTE, Noakhali Science and Technology University, Noakhali, Bangladesh
Arup Mazumder, Department of Electronics and Telecommunication Engineering, Southeast University, Dhaka, Bangladesh*

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

22. PaperID 30111740: Risk Management Process Analysis for Information & Communication Technology (ICT) Systems: Risk Mitigation Perspective (pp. 170-175)

(1) Nwagu, Chikezie Kenneth; (2) Omankwu, Obinnaya Chinecherem; (3) Okonkwo, Obikwelu Raphael

- (1) *Computer Science Department, Nnamdi Azikiwe University Awka Anambra State, Nigeria.*
(2) *Computer Science Department, Michael Okpara University of Agriculture Umudike Umuahia, Abia State, Nigeria.*
(3) *Computer Science Department, Nnamdi Azikiwe University Awka Anambra State, Nigeria.*

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

23. PaperID 30111741: Credit Card Fraud Detection Using Neural Network (pp. 176-180)

- (1) *Nwagu, Chikezie Kenneth; (2) Omankwu, Obinnaya Chinecherem; (3) Inyama, Hycient*
(1) *Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,*
(2) *Computer Science Department, Michael Okpara University of Agriculture, Umudike Umuahia, Abia State, Nigeria*
(3) *Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria*

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

24. PaperID 30111742: Reinforcement Learning a Tool for Filtering Personalized Web Document (pp. 181-185)

- (1) *Nwagu, Chikezie Kenneth; (2) Omankwu, Obinnaya Chinecherem; (3) Inyama, Hycient*
(1) *Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,*
(2) *Computer Science Department, Michael Okpara University of Agriculture, Umudike Umuahia, Abia State, Nigeria*
(3) *Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria*

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

25. PaperID 30111744: Influence of Distributed Generation on the Protection Scheme of Local Distribution System (pp. 186-190)

Usman Yousaf, Dr. Azzam ul Asar, Alishpa Khattak
Department of Electrical Engineering, CECOS University, Peshawar, Pakistan

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

26. PaperID 30111747: Sentiment Analysis of Moroccan Tweets using Naive Bayes Algorithm (pp. 191-200)

Abdeljalil ELABDOULI, Larbi HASSOUNI, Houda ANOUN
RITM Laboratory, CED Engineering Sciences, Ecole Supérieure de Technologie, Hassan II University of Casablanca, Morocco

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

27. PaperID 30111749: Using Biometric Watermarking for Video File Protection Based on Chaotic Principle (pp. 201-206)

Saja J. Mohammed, Computer Science Department, College of Computer Science & Mathematics, Mosul University Mosul, IRAQ

Full Text: [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

28. PaperID 30111752: Dynamic Cluster Formation Method to Increase the Lifetime of a Wireless Sensor Network (pp. 207-215)

*K. Johny Elma, Assistant Professor, Department of IT, Jeppiaar SRR Engineering College, Chennai, India
Dr. S. Meenakshi, Professor and Head, Department of IT, Jeppiaar SRR Engineering College, Chennai, India*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

29. PaperID 30111757: Context based Power Aware Multi-Effector Action optimized Reinforcement Learning (pp. 216-224)

Mr. S. Senthil Kumar, Research Scholar. & Dr. T. N. Ravi, Assistant Professor, Periyar EVR College

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

30. PaperID 30111759: Comparative Analysis of Convergence Times between OSPF, EIGRP, IS-IS and BGP Routing Protocols in A Network (pp. 225-227)

*Eunice Domfeh Asabere, Joseph Kobina Panford, James Ben Hayfron-Acquah
Dept. Of Computer Science, KNUST, Kumasi, Ghana*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

31. PaperID 30111760: Arabic Spelling Checker Algorithm for Speech Recognition (pp. 228-235)

Huda F. Al-shahad, Department of Computer Science, University of Kerbala, Karbala, Iraq

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

32. PaperID 30111761: Detection of Physiological Changes in Women during Muscle Activity Using Electromyography and Other Techniques (pp. 236-241)

Rossana Rodríguez Montero (a), David Asael Gutiérrez Hernández (a), Miguel Mora González (b), Víctor Manuel Zamudio Rodríguez (a), Juan Martín Carpio Valadez (a), Manuel Ornelas Rodríguez (a), Claudia Margarita Lara Rendón (a), Miguel Salvador Gómez Díaz (a), Jesús Ricardo Sevilla Escoboza (b), and Víctor Porfirio Vera Ávila (b)

(a). Tecnológico Nacional de México. Instituto Tecnológico de León. León, Guanajuato, México.

(b). Universidad de Guadalajara. Centro Universitario de los Lagos. Lagos de Moreno, Jalisco, México.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

33. PaperID 30111768: Prostate Near-Infrared Fluorescent Image Segmentation Using Hopfield Neural Network Classifier (pp. 242-251)

Rachid Said Sammouda, Computer Science Department, College of Computer and Information Sciences, King Saud University, Riyadh, KSA

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

34. PaperID 30111772: New Image Processing Techniques Using Elitism Immigrants Multiple Objective of Genetic Algorithms for Disease Detection (pp. 252-260)

Dr. Khalil Ibrahim Mohammed Abuzanounneh, IT Department, College of Computer, Qassim University, KSA

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

35. PaperID 30111774: Reliability of Scalable Coherent Interface in Cluster Interconnection (pp. 261-265)

Mr Adinew Belay, HOD, Department of CS, SCI, Mizan Tepi University, Eithiopia.

Dr Narasimha Rao Yamarthi, Professor, Department of CS, SCI, Mizan Tepi University, Eithiopia.

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

36. PaperID 30111777: Effort Prediction Tool using UML Diagrams (EPT) (pp. 266-274)

Atica M. Altaie, Asma Yassin Hamo

Software Engineering Department, Computer and Mathematics Science College, University of Mosul, Mosul, Iraq

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

37. PaperID 30111778: Design and Implementation Elliptic Curve Digital Signature Algorithm using Multi Agent System (pp. 275-282)

Tawfeeq M. Tawfeeq Al-Flaih, Software Engineering Department, College of Computer Science and Mathematics University of Mosul, Mosul, IRAQ

Marwa Adeeb Al-jawaherry, Software Engineering Department, College of Computer Science and Mathematics University of Mosul, Mosul, IRAQ

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

38. PaperID 30111779: Firefly Algorithm Implementation Based on Arduino Microcontroller (pp. 283-288)

Riyadh Zaghlool Mahmood, Software Engineering Department, College of Computer Science and Mathematics, University of Mosul, Mosul, IRAQ

Marwa Adeeb Al-jawaherry, Software Engineering Department, College of Computer Science and Mathematics, University of Mosul, Mosul, IRAQ

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

39. PaperID 30111781: Analysis of Android Bugs for Mobile Applications (pp. 289-293)

Tawfeeq M. Tawfeeq Al-Flaih, Software Engineering Department, College of Computer Science and Mathematics, University of Mosul, Mosul, IRAQ

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

40. PaperID 30111771: An Enhanced Pipeline for Improving Reliability using Evolution and Optimization Technique (pp. 294-307)

Dr Narasimha Rao Yamarthi (1), Mr Adinew Belay (2), Mr Abdisa Lechisa (3), P.Narasimha Rao (4).
(1) Professor, Department of Computer Science, School of Computing & Informatics, Mizan Tepi University, Ethiopia.

(2) HOD, Department of Computer Science, School of Computing & Informatics, Mizan Tepi University, Ethiopia.

(3) Lecturer, Department of Computer Science, School of Computing & Informatics, Mizan Tepi University, Ethiopia.

(4) Lecturer, Department of Computer Science, School of Computing & Informatics, Mizan Tepi University, Ethiopia.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

41. PaperID 30111773: An Efficient Voice Based Person Identification System for Secured Communication (pp. 308-325)

Dr. Shakila Basheer, Lecturer in Information system Department, King Khalid University, ABHA, Saudi Arabia

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

42. PaperID 30111766: Comparative Analysis of Modern Methods and Algorithms of Cryptographic Protection of Information (pp. 326-330)

Saleh I. Alomar & Saleh A. Khawatreh

AL-Ahliyya Amman University, Department of Engineering

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

43. PaperID 30111775: SQS: An Ontology Based Framework for Evaluating Service Oriented Software Applications: A case study of E-Governance Services (pp. 331-336)

Ezekiel U, Okike, University of Botswana, Gaborone, Botswana

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

Risk Management Process Analysis for Information & Communication Technology (ICT) Systems: Risk Assessment perspective

Nwagu, Chikezie Kenneth, Omankwu, Obinnaya Chinecherem, Okonkwo, Obikwelu Raphael

¹**Computer Science Department, Nnamdi Azikiwe University Awka Anambra State, Nigeria.
Nwaguchikeziekeneth@hotmail.com**

²**Computer Science Department, Michael Okpara University of Agriculture Umudike Umuahia, Abia State,
Nigeria.
Saintbeloved@yahoo.com**

³**Computer Science Department, Nnamdi Azikiwe University Awka Anambra State, Nigeria.
Oobi2971@yahoo.com**

Abstract— *This Research Paper, explored Risk management from three perspectives: Risk Assessment, Risk Mitigation & Continuous Risk monitor, evaluate and Review. Risk Assessment Perspective highlights vulnerability/threat sources, types of Security, methods used to “break” security. The assessment processes were fully depicted showing its specific steps. It also reflected importance of consistent communication and consultation among the stakeholders as it is key and sustained throughout the entire processes. This work also strengthens the fact that risk is not entirely negative as reveals by its SWOT (Strength, Weakness, Threat and Opportunity) analysis. Hence one can build on the opportunity thereby reducing the associated threat and also improve on the strengths. This invariably reduces and if possible eradicates the weaknesses.*

Keywords—*Risk, Risk Management, Risk Assessment, Vulnerability, threat, Opportunity*

I. INTRODUCTION

In the face of global ever-dynamic threats and attacks, every Organization is adopting measures to reduce negative risks and utilize positive risks. This ensures that her vision and mission is protected, guarded and fully enhanced. This critical as Organizations make ICT, hub for better support and sustenance of her business. As Organizations automated their processes using data and communication devices, Risk Management plays a very critical role in protecting the organizations information assets, and therefore its mission. Risk Management is every stakeholder`s duty and not only for the technical IT team. Therefore, it should be treated as fundamentally as an essential role of the Management. An effective and efficient risk management process is an important component of a successful ICT security so as to ensure data confidentiality, integrity and high availability. **Vulnerability** is the devices` weakness which can be accidentally triggered or intentionally exploited. While **opportunity** is just positive risk which can be invested upon in order to maximize the use and benefit the devices. Based on this, Organizations are continuously working on reducing the vulnerabilities by minimizing sources of threats and maximizing the opportunities by strengthening the securities during **SWOT (Strength-weakness-Opportunities and threats)** analysis. **Risk management** is the process of identifying risk, assessing risk, and taking steps to reduce risk to an acceptable level, if possible eradicate it completely. The objectives of Risk assessment in this work are to increase the likelihood and impact of positive events, and decrease the likelihood of negative events in devices. However, before now, Risk management is not consciously or transparently carried out for data and communication devices as the “practice of the day” is that the Organization`s perceived final step in the system development life cycle of the devices is always junking of the devices without final proper risk assessment to ensure that no critical piece of information or data can be intentional or accidentally exploited.

II. VULNERABILITIES/THREAT SOURCES

Risks are continuously evolving. The goal of this step is to identify the potential vulnerability and threat-sources that are applicable to the ICT systems being evaluated. A threat-source is defined as any circumstance or event with the potential to cause harm to an ICT system. The common threat-sources can be natural, human, or environmental:

- Natural Threats: Floods, earthquakes, tornadoes, landslides, avalanches, electrical storms, and other such events.
- Human Threats: Events that are either enabled by or caused by human beings, such as unintentional acts (inadvertent data entry) or deliberate actions (network based attacks, malicious software upload, unauthorized access to confidential information).
- Environmental Threats: Long-term power failure, pollution, chemicals, liquid leakage.

In assessing threat-sources, it is important to consider all potential threat-sources that could cause harm to an ICT System and its processing environment. For example, although the threat statement for an ICT System located in a desert may not include natural flood, because of the low likelihood of such an events occurring, environmental threats such as a bursting pipe can quickly flood a computer room and cause damage to an organization's IT assets and resources.

Humans can be threat-sources through intentional acts, such as deliberate attacks by malicious persons or disgruntled employees, or unintentional acts, such as negligence and errors.

A deliberate attack can be either:

- A malicious attempt to gain unauthorized access to an ICT System (e.g., via password guessing) in order to compromise system and data integrity, availability, or confidentiality or
- A benign, but nonetheless purposeful, attempt to circumvent system security.

Motivation and Threat Actions

According to Siciliano, (2011) Hackers are motivated by a number of factors such as ego, religion, politics, activism etc. Motivation and the resources for carrying out an attack make humans potentially dangerous threat-sources. In addition, reviews of the history of system break-ins; security violation reports; incident reports; and interviews with the system administrators, help desk personnel, and user community during information gathering help identify human threat-sources that have the potential to harm an IT system and its data and that may be a concern where vulnerability exists.

With these information, organizations should be mindful of them and consciously have proofs against them to reduce and/or prevent successful exploits.

Types of Security

Secure communication is when two or more devices are communicating without eavesdropping or interception by a third party. This communication involves sharing of data and information with varying confidentiality and integrity. Among the means to achieve this is:

- Code: This is a means whereby the content and nature of communication is hidden. It is a rule to convert a piece of information and data (for example, a letter, word, phrase or gesture) into another form of representation, not necessarily of same type.
- Encryption: This is also another means whereby the nature and content of communication is hidden. Here, data and communication is rendered hard to read to any unauthorized party. In some highly security-conscious environments, encryption is configured such that it is a basic requirement for connection and communication to be established. No room for opportunistic encryption which is a lower security method to generally increase percentage of generic traffic and this makes the content susceptible to eavesdropping.

- Steganography: This is sometimes referred to as “hidden writing” in which data can be hidden within another, mostly innocuous data. In this way, it is difficult to find or remove unless you know how to find it. For example in communication, the hiding of important data such as telephone number in apparently innocuous data (an MP3 music file). A good advantage of this is plausible deniability – unless one can prove that the data is there (which is usually not easy), it is deniable that the file contains any.
- Identity based Networks: Unwanted or malicious behaviour is possible on the web since it is inherently anonymous. Identity based network removes the chance of anonymity as the identity of the sender and recipient are known.
- “Security by Obscurity: Similar to needle in a haystack in which secrecy of design or implementation is used to provide security. Though this is discouraged and not recommended by standard bodies. But stakeholders believe that if the flaws are not known, then attackers will be unlikely to find them. As it is known that attacker’s first step is usually information gathering which is delayed by this.
- Random Traffic: This involves creating random data flow to make the presence of genuine communication harder to detect and traffic analysis less reliable.
- Hard to trace routing methods: This method hides the parties involved in a communication through unauthorized third-party systems or relays.

Methods used to “break” security

- Bugging: This is simply known as covert listening device which involves miniature transmitter and microphone. This enables unauthorized parties to listen to conversation.
- Computers (general): Any security obtained from a computer is limited by the many ways it can be compromised – by hacking, keystroke logging, backdoors or even in extreme cases by monitoring the tiny electrical signals given off by keyboard or monitors to reconstruct what is typed or seen.
- Laser audio Surveillance: Sounds including speech inside rooms can be sensed by bouncing a laser beam off a window of the room where a conversation is held and detecting and decoding the vibrations in the glass caused by the sound waves.
- Spoofing: This is a situation in which one person or program successfully masquerades as another by falsifying data and thereby gaining an illegitimate advantage or access. For example, Caller Id, Email address, IP address etc. can all be spoofed.

III. RISK ASSESSMENT PROCESSES

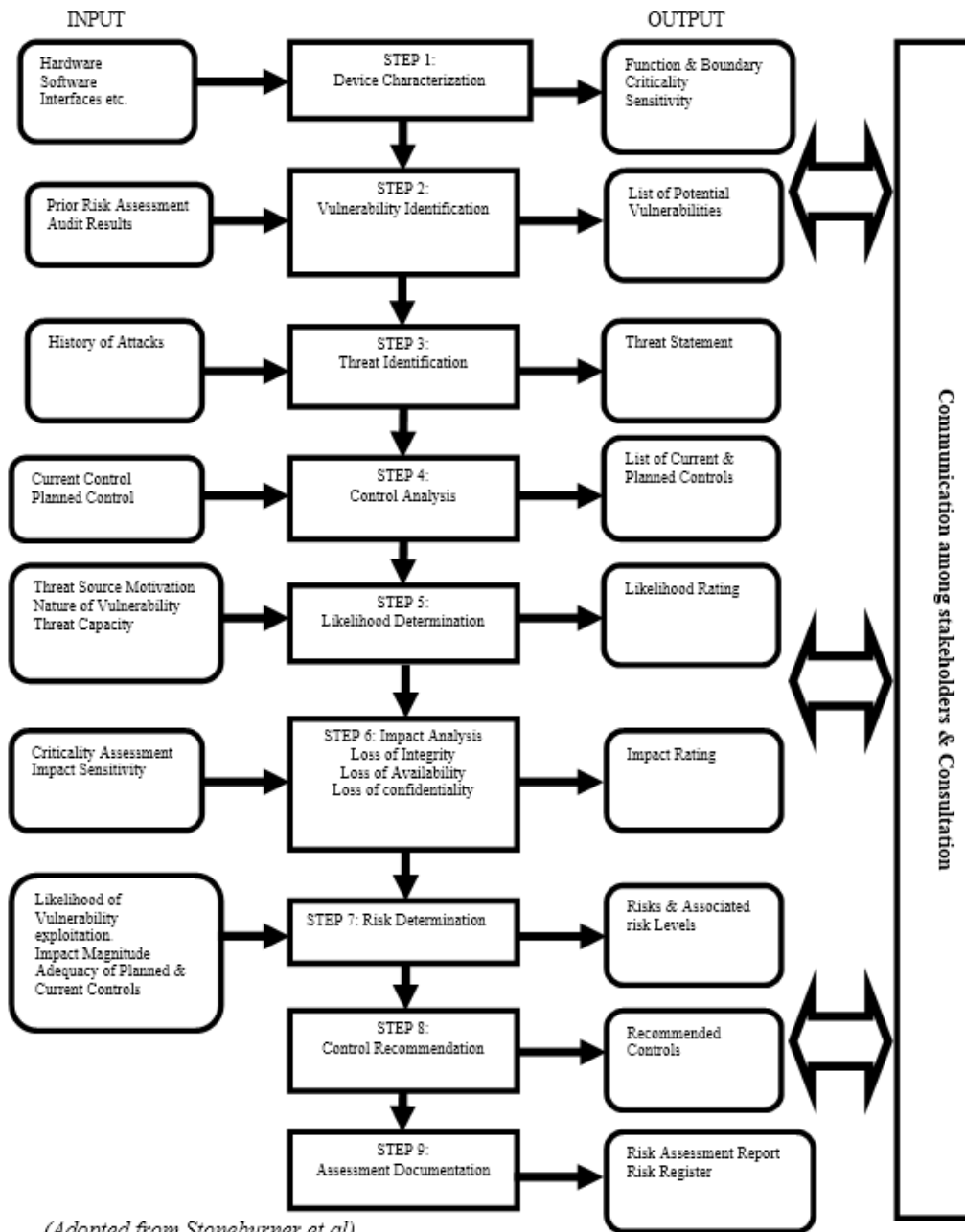
Risk Management involves three main principal processes: risk assessment, risk Mitigation and, Monitor, evaluate and review. Risk Assessment has many steps (see figure 1) with processes as broadly summarized:

- Risk identification: This allows individuals to identify risks so that the stakeholders will be in the know of potential problems inherent in the devices. It is pertinent to start this stage as early as possible and should be repeated frequently.
- Risk analysis and Priority: Risk analysis transforms the estimates or data about specific risks that developed during risk identification into a consistent form that can be used to make decisions around prioritization. Risk prioritization enables operations to commit resources to manage the most important risks.
- Risk register (Statements) integration: This is the result of risk assessment process. It is a document which contains lists of identified risks, root causes of risks, lists of potential responses, risk owners, symptoms and

warning signs, relative rating or priority list. Risk for additional analysis and responses, and a watch list which is a list of low-priority risk within the risk register.

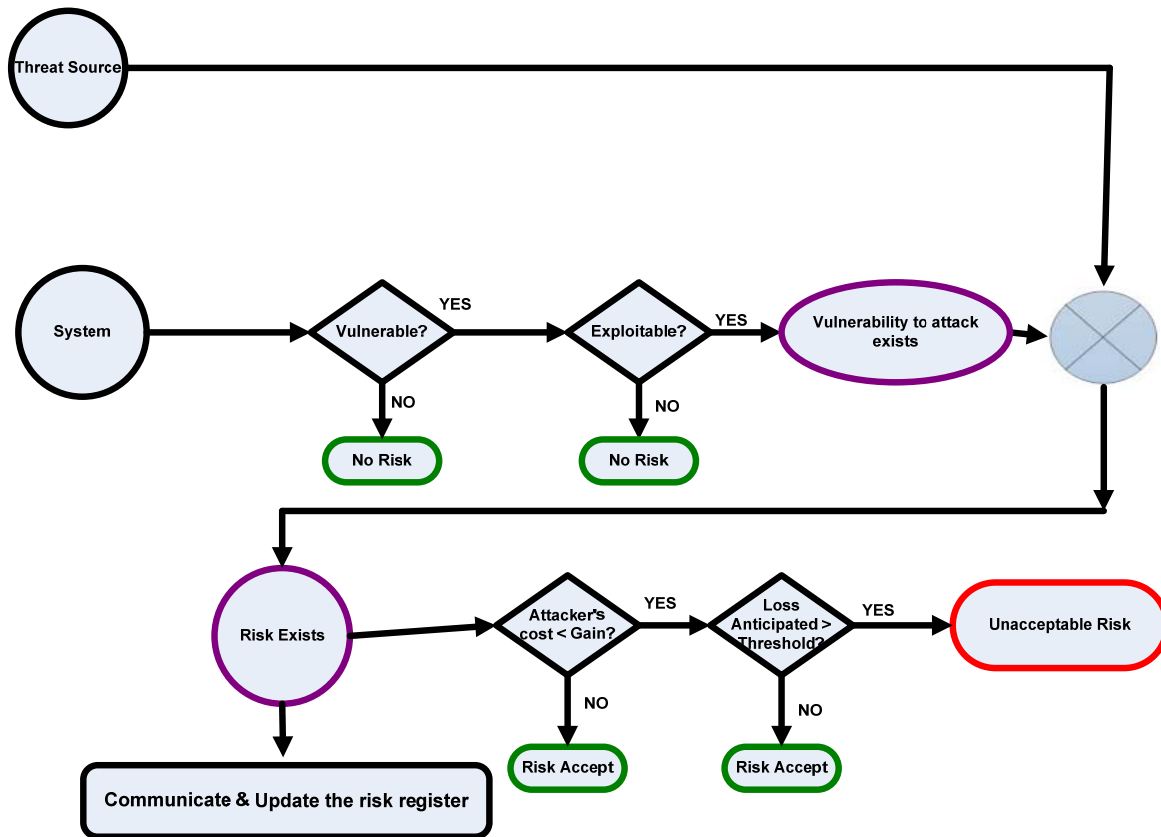
- Consistent Communication and Consultation: There is steady communication among stakeholders within the organization as everyone is practically involved. In addition to this, the stakeholders can consult the manufacturer of the device through any of the appropriate channels such as through their Representative or customer voices. This ensures speedy and reliable Response.

Figure I : RISK ASSESSMENT PROCESSES CYCLE



The processes are followed up with consistent updates and awareness campaigns among stakeholders as new challenges, discoveries and prospects arise.

Figure II: SIMPLIFY OVERALL RISK ASSESSMENT PROCESS FLOWCHAT FOR ICT SYSTEM



(Adopted from Stoneburner et al)

Principal stakeholders, knowing the potential risks and recommended controls, may ask, “When and under what circumstances should I take action? When shall I implement these controls to mitigate the risk and protect our organization?” Below system architecture is the answer. The system architecture is further articulated in the following rules of thumb, which provide guidance on actions to mitigate risks from intentional human threats:

- When vulnerability (or flaw, weakness) exists → implement initial risk assessment to reduce the likelihood of a vulnerability’s being exercised.
- When vulnerability can be exercised → apply layered protections, architectural designs, and administrative controls to minimize the risk or prevent the occurrence.
- When the attacker’s cost is less than the potential gain → apply protections to decrease an attacker’s motivation by increasing the attacker’s cost (e.g., use of system controls such as limiting what a system user can access and do can significantly reduce an attacker’s gain), communicate and update the risk registers.
- When loss is too great → apply design principles, layered architectural designs, and technical and nontechnical protections to limit the extent of the attack, thereby reducing the potential for loss, update the risk register and communicate.

IV. CONCLUSIONS

One of the main aims of Risk Management process analysis for ICT systems: Risk Assessment perspective is to ensure that the devices are deployed with best of security measures in place, which makes the organization to be fully proactive rather than reactive as it is today with many organizations. This will, invariably increase stakeholders' risk appetite for positive risks and of course, establish a careful risk threshold for negative risks. This further ensures that cost of attack from a potential intentional attacker is far higher than the anticipated gain which would likely discourage the attackers. A successful attack has high currency impact, loss of customer confidence and negative business reputation. It assists management to make well-informed risk management decisions to justify huge capital expenditures that are part of an ICT budget and also in authorizing the ICT devices on the basis of the supporting documentation resulting from the performance of risk management. With the flowchart, stakeholders are convinced of what to do, hence proactively take the right steps/decision to protect the organization and ensure optimal utilization of the ICT systems.

It is worthy of note here that the process is a continuous one in order to get optimal throughput from the devices with little or no down time as a result of attack, vulnerability or negative risks. Thereby increasing the opportunities which the devices can offer. Therefore Risk management continues even at the final stage of systems development Life cycle which is disposal of the devices. It is pertinent to carry out risk assessment at the disposal stage to ensure sensitive data or information are not left out such as vital configurations which may include plain-text passwords, administrator credentials etc.

V REFERENCES

Stoneburner G.,Goguen A. and Feringa A.(2001,July). Risk Management Guide for information Systems. Retrieved September 4,2014 from <http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf>.

Radack S.(Ed.).Security for wireless network and Devices. Retrieved August 20,2014 from <http://www.itl.nist.gov/lab/bulletns/bltnmar03.htm>

Cooney M.(2012,September 21).10 common mobile Security problems to attack. Retrieved August 22,2014 from <http://www.pcworld.com/article/2010278/10-common-mobile-security-problems-to-attack.html>

A Guide to the Project Management Body of Knowledge(PMBOK) Fifth Edition.

Alshboul A.(2010).Information Systems Security measures and countermeasures: Protecting Organizational Assets from malicious Attacks,3,Article 486878.Retrieved September 10,2014 from <http://www.ibimapublishing.com/journals/CIBIMA/2010/486878/486878.pdf>

Secure Communication. Retrieved September 9,2014 from http://en.wikipedia.org/wiki/Secure_communication.

http://www.labcompliance.com/tutorial/risk/default.aspx?sm=d_a. Accessed September 9,2014

<http://www.iip.kit.edu/english/1680.php> Accessed September 7,2014

Risk Management Process Overview. Accessed September 3,2014 from <http://technet.microsoft.com/en-us/library/cc535304.aspx>

Schneier, B. (2008). Security through Obscurity.

Siciliano, R. S. (2011). Seven Types of Hacker Motivations. Infosec Island

Evolutionary Perspective of Robotics

Nwagu, Chikezie Kenneth, Omankwu, Obinnaya Chinecherem, and Inyama, Hycient

¹ Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,
Nwaguchikeziekeneth@hotmail.com

² Computer Science Department, Michael Okpara University of Agriculture, Umudike
Umuahia, Abia State, Nigeria
saintbeloved@yahoo.com

³ Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka
Anambra State, Nigeria.

ABSTRACT

The evolution of application fields and their sophistication have influenced research topics in the robotics community. This evolution has been dominated by human necessities. In the early 1960s, the industrial revolution put industrial robots in the factory to release the human operator from risky and harmful tasks. The later incorporation of industrial robots into other types of production processes added new requirements that called for more flexibility and intelligence in industrial robots.

Keywords: Robotics, artificial intelligence, algorithm.

Introduction

During the last 45 years, robotics research has been aimed at finding solutions to the technical necessities of applied robotics. The evolution of application fields and their sophistication have influenced research topics in the robotics community. This evolution has been dominated by human necessities. In the early 1960s, the industrial revolution put industrial robots in the factory to release the human operator from risky and harmful tasks. The later incorporation of industrial robots into other types of production processes added new requirements that called for more flexibility and intelligence in industrial robots. Currently, the creation of new needs and markets outside the traditional manufacturing robotic market (i.e., cleaning, demining, construction, shipbuilding, agriculture) and the aging world we live in is demanding field and service robots to attend to the new market and to human social needs.

This article addresses the evolution of robotics research in three different areas: robot manipulators, mobile robots, and biologically inspired robots. Although these three areas share some research topics, they differ significantly in most research topics and in their application fields. For this reason, they have been treated separately in this survey. The section on robot manipulators includes research on industrial robots, medical robots and rehabilitation robots, and briefly surveys other service applications such as refueling, picking and palletizing. When surveying the research in mobile robots we consider terrestrial and underwater vehicles. Aerial vehicles are less widespread and for this reason have not been considered. Biologically inspired robots include mainly walking robots and humanoid robots; however, some other biologically inspired underwater systems are briefly mentioned. In spite of the differences between robot manipulators, mobile robots and biologically inspired robots, the three research areas converge in their current and future intended use: field and service robotics. With

the modernization of the First World, new services are being demanded that are shifting how we think of robots from the industrial viewpoint to the social and personal viewpoint. Society demands new robots designed to assist and serve the human being, and this harks back to the first origins of the concept of the robot, as transmitted by science fiction since the early 1920s: the robot as a human servant (see Figure 1). Also, the creation of new needs and markets outside the traditional market of manufacturing robotics leads to a new concept of robot. A new sector is therefore arising from robotics, a sector with a great future giving service to the human being. Traditional industrial robots and mobile robots are being modified to address this new market. Research has evolved to find solutions to the technical necessities of each stage in the development of service robots.



Figure 1. ASIMO. Photograph courtesy of American Honda Motor Co.

Robot Manipulators

A robot manipulator, also known as a robot arm, is a serial chain of rigid limbs designed to perform a task with its end-effector. Early designs concentrated on industrial manipulators, to perform tasks such as welding, painting, and palletizing. The evolution of the technical necessities of society and the technological advances achieved have helped the strong growth of new applications in recent years, such as surgery assistance, rehabilitation, automatic refuelling, etc. This section surveys those areas that have received a special, concentrated research effort, namely, industrial robots, medical robots, and rehabilitation robots.

Industrial Robots

It was around 1960 when industrial robots were first introduced in the production process, and until the 1990s industrial robots dominated robotics research. In the beginning, the automotive industry dictated the specifications industrial robots had to meet, mainly due to the industry's market clout and clear technical necessities. These necessities determined which areas of investigation were predominant during that period.

One such area was kinematic calibration, which is a necessary process due to the inaccuracy of kinematic models based on manufacturing parameters. The calibration process is carried out in four stages. The first stage is mathematical modeling, where the Denavit-Hartenberg (DH) method and the product-of-exponential (POE) formulation lead the large family of methods. A detailed discussion of the fundamentals of kinematic modeling can be found in the literature [1]. The gap between the theoretical model and the real model is found in the second stage by direct measurement through sensors. Thus, the true position of the robot's end effector is determined, and by means of optimization techniques, the parameters that vary from their nominal values are identified in the third stage. Last, implementation in the robot is the process of incorporating the improved kinematic model. This process will depend on the complexity of the machine, and iterative methods will have to be employed in the most complex cases. Research in robot calibration remains an open issue, and new methods that reduce the computational complexity of the calibration process are still being proposed [2], [3].

Another important research topic is motion planning, wherein subgoals are calculated to control the completion of the robot's task. In the literature there are two types of algorithms, implicit methods and explicit methods. Implicit methods specify the desired dynamic behavior of the robot. One implicit scheme that is attractive from the computational point of view is the potential field algorithm [4]. One disadvantage of this approach is that local minima of the potential field function can trap the robot far from its goal. Explicit methods provide the trajectory of the robot between the initial and final goal. Discrete explicit methods focus on finding discrete collision-free configurations between the start and goal configurations. These methods consist mainly of two classes of algorithms, the family of road-map methods that include the visibility graph, the Voronoi diagram, the free-way method and the Roadmap algorithm [5], and the cell-decomposition methods [6]. Continuous explicit methods, on the other hand, consist in basically open-loop control laws. One important family of methods is based on optimal-control strategies [7], whose main disadvantages are their computational cost and dependence on the accuracy of the robot's dynamic model.

Besides planning robot motion, control laws that assure the execution of the plan are required in order to accomplish the robot's task. Thus, one fundamental research topic focuses on control techniques. A robot manipulator is a nonlinear, multi-variable system and a wide spectrum of control techniques can be experimented here, ranging from the simpler proportional derivative (PD) and proportional integral derivative (PID) control to the computed-torque method [8], and the more sophisticated adaptive control [9] whose details are out of the scope of this survey.

Typical industrial robots are designed to manipulate objects and interact with their environment, mainly during tasks such as polishing, milling, assembling, etc. In the control of the interaction between manipulator and environment, the contact force at the manipulator's end effector is regulated. There are diverse schemes of active force control, such as stiffness control, compliant control, impedance control, explicit force control and hybrid force/position control.

The first three schemes belong to the category of indirect force control, which achieves force control via motion control, while the last two methods perform direct force control by means of explicit closure of the force-feedback loop. Readers who wish to study this subject in detail will find an interesting account in [10].

An attractive alternative for implementing force-control laws is the use of passive mechanical devices so that the trajectory of the robot is modified by interaction forces due to the robot's own accommodation. An important example of passive force control is the remote center of compliance (RCC) system patented by Watson in 1978 [11] for peg-in-hole assembly. Passive force control is simpler than active force control laws but has disadvantages, such as lacking flexibility and being unable to avoid the appearance of high contact forces.

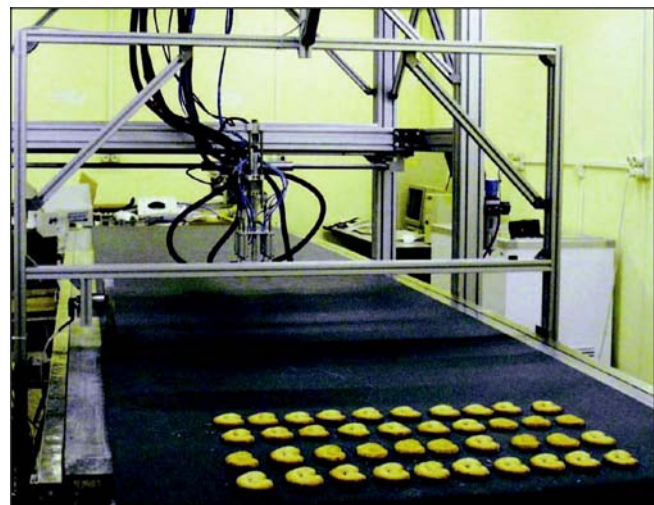


Figure 2. Robots in the food industry.

As 1990 began, new application areas for industrial robots arose that imposed new specifications, with flexibility as the principal characteristic. The new industries that introduced industrial robots in their productive process were the food and pharmacy industries (see Figure 2). Postal services too looked for robotic systems to automate their logistics. The main requirement was the capacity to accommodate variations in product, size, shape, rigidity (in the case of foods), etc. The ability to self-adapt to the product and the environment became the issue in the following lines of investigation in the area of industrial robotics. The main line of research now is aimed at equipping the control system with sufficient intelligence and problem-solving capability. This is obtained by resorting to artificial-intelligence techniques. Different artificial intelligence (AI) techniques are used to provide the robot with intelligence and flexibility so it can operate in dynamic environments and in the presence of uncertainty. Those techniques belong to three areas of artificial intelligence: learning, reasoning and problem solving [12]. Among the diverse learning algorithms, inductive learning is the most widely used in robotics, in which the robot learns from preselected examples

Typical reasoning paradigms in robotics include fuzzy reasoning [14], mostly used in planning under uncertainty, spatial reasoning, and temporal reasoning. The techniques most commonly used in robotics for problem solving are means-end reasoning, heuristic searching, and the blackboard (BB) model.

Another solution to the control of robots in dynamic or unknown environments consists of introducing the operator in the control loop, such that the robot is remotely operated. The success of a teleoperation system relies on the correct feedback of the robot interaction with the environment, which can be visual, tactile or force reflection. The greatest disadvantage that teleoperated systems involve are transmission delays when the distance between the operator and the robot is significant, like in space teleoperation or over the Internet. Some research has explored solutions to this modified to respond to this new market, yielding surgery robots, refueling robots, picking and palletising robots, feeding robots, rehabilitation robots, etc. Two of the most relevant service applications of robot manipulators are in the field of medical robots and rehabilitation robots that are catching the interest of researchers all over the world. In the following subsections, we will summarize research topics in medical robotics and rehabilitation robotics.

Medical Robots

In recent years, the field of medicine has been also invaded by robots, not to replace qualified personnel such as doctors and nurses, but to assist them in routine work and precision tasks. Medical robotics is a promising field that really took off in the 1990s. Since then, a wide variety of medical applications have emerged: laboratory robots, telesurgery, surgical training, remote surgery, telemedicine and teleconsultation, rehabilitation, help for the deaf and the blind, and hospital robots. Medical robots assist in operations on heart-attack victims and make possible the millimeter-fine adjustment of prostheses. There are, however, many challenges in the widespread implementation of robotics in the medical field, mainly due to issues such as safety, precision, cost and reluctance to accept this technology.

Medical robots may be classified in many ways: by manipulator design (e.g., kinematics, actuation); by level of autonomy (e.g., preprogrammed versus teleoperation versus constrained cooperative control); by targeted anatomy or technique (e.g., cardiac, intravascular, percutaneous, laparoscopic, micro-surgical); by intended operating environment [e.g., in-scanner, conventional operating room (OR)], etc. Research remains open in the field of surgical robotics, where extensive effort has been invested and results are impressive. Some of the key technical barriers include safety [16], where some of the basic principles at issue are redundancy, avoiding unnecessary speed or power in actuators, rigorous design analysis and multiple emergency stop and checkpoint/restart facilities. Medical human-machine interfaces are another key issue that draws upon essentially the same technologies as other application domains.

problem, such as interposing a virtual robot in charge of environment feedback, but this procedure is only valid if the robot works in structured environments. Another solution is teleprogramming, in which the operator sends high-level commands and the robot carries out the task in closed-loop control. Recently, considerable attention has been devoted to Internet-based teleoperation, in which the transmission delay is variable. For direct force feedback, wave-variable-based approaches have been used extensively, and they have been further extended to include estimation and prediction of the delay. A comprehensive survey can be found in [15].

With the rapid modernization of the First World, new types of services are being required to maintain a certain quality of life. A new, promising robotics sector is arising to serve the human being. Traditional industrial robots are being

Surgeons rely on vision as their dominant source of feedback; however, due to the limited resolution of current-generation video cameras, there is interest in optical overlay methods, in which graphic information is superimposed on the surgeon's field of view to improve the information provided [17]. As surgeons frequently have their hands busy, there has been also interest in using voice as an interface. Force and haptic feedback is another powerful interface for telesurgery applications [18]. Much of the past and present work on telesurgery involves the use of master-slave manipulator systems [19], [20]. These systems have the ability to feed forces back to the surgeon through the master manipulator, although slaves' limitations in sensing tool-to-tissue forces can somewhat reduce this ability.

The field of medical robotics is expanding rapidly and results are impressive as a large number of commercial devices are being used in hospitals. However, societal barriers have to be overcome and significant engineering research effort is required before medical robots have wide-spread impact on health care.

Rehabilitation Robots

Activity in the field of rehabilitation robotics began in the 1960s [21] and has slowly evolved through the years to a point where the first commercially successful products are now available. Today, the concept of "rehabilitation robot" may include a wide array of mechatronic devices ranging from artificial limbs to robots for supporting rehabilitation therapy or for providing personal assistance in hospital and residential sites. Examples include robots for neuro-rehabilitation [22], power-augmentation orthosis [23], rehabilitative orthosis, etc. The field of rehabilitation robotics is less developed than that of industrial robotics. Many assistive robotic systems have featured an industrial robot arm for reasons of economy and availability [24]. However, the specifications for robots in these two application areas are very different. The differences arise from the involvement of the user in rehabilitation applications. Industrial robots are typically powerful and rigid to provide speed and accuracy. They operate autonomously and, for reasons of safety, no human interaction is permitted.

Rehabilitation robots must operate more slowly and be more compliant to facilitate safe user interaction. Thus, rehabilitation robotics is more akin to service robotics, which integrates humans and robots in the same task. It requires safety and special attention must be paid to human-machine interfaces that have to be adapted for disabled or nonskilled people operating a specific programming device. It is also recognized that there is a need for research and development in robotics to focus on developing more flexible systems for use in unstructured environments. The leading developments of this type in rehabilitation robotics concern, among other topics, mechanical design (including mobility and end-effectors), programming, control and man machine interfaces [25]. Subsection "Humanoid Robots" of this article expands on new research into human-robot interaction.

Mobile Robots

The term mobile robot describes a robotic system able to carry out tasks in different places and consisting of a platform moved by locomotive elements. The choice of the locomotive system depends firstly on the environment in which the robot will operate. This can be aerial, aquatic or terrestrial. In the aquatic and aerial environments, the locomotive systems are usually propellers or screws, although at the seabed legs are also used. The choice of the locomotive system on earth is more complicated due to the variety of terrestrial environments. Wheels, tracks, and legs are typical terrestrial locomotive elements.

Mobility provides robots with enhanced operating capacity and opens up new areas of investigation. Some such areas are common to all mobile robots, like the navigation problem, whereas others deal more specifically with a certain locomotion system, like the walking gait.

Practically by the time industrial robots were introduced in the production process, mobile robots were installed in the factory. This was around 1968, and the robots were mainly automated guided vehicles (AGVs), vehicles transporting tools and following a predefined trajectory. Nevertheless, the research in this area deals now with autonomous indoor and outdoor navigation. Autonomous mobile-robot navigation consists of four stages: perception of the environment, self-localization, motion planning and motion generation.

In structured environments, the perception process allows maps or models of the world to be generated that are used for robot localization and motion planning. In unstructured or dynamic environments, however, the robot has to learn how to navigate. Navigation is, therefore, one of the main applications of artificial intelligence to robotics, where learning, reasoning and problem solving come together. The main research in mobile robotics is focusing on robot localization and map generation.

Conclusion

Since the introduction of industrial robots in the automotive industry, robotics research has evolved over time towards the development of robotic systems to help the human in dangerous, risky or unpleasant tasks. As the complexity of tasks has increased, flexibility has been

demanding in industrial robots, and robotics research has veered towards adaptive and intelligent systems.

Since 1995, robotics research has entered the field- and service-robotics world, where we can find manipulators, mobile robots and animal-like robots with great perspectives of development and increasing research interest. Surgical robots have been the first successes, and recently different areas in medical-and rehabilitation-robotics applications have arisen. Other examples can be found in the fields of home cleaning, refueling and museum exhibitions, to name just a few areas.

Service-robotics research is also aimed at providing a comfortable, easy life for the human being in an aging world. The United Nations Economic Commission for Europe (UNECE) forecasts strong growth of professional robots in application areas such as humanoid robots, field robots, underwater systems and mobile robot platforms for multiple use in the period of 2005–2008 [86]. The UNECE also forecasts a tremendous rise in personal robots in the next few years. Robotics research has to make a great effort to solve in very few years the challenges of this new field of research, which will be largely determined by interaction between humans and robots. Figure 10 summarizes the evolution of robotics research over the last 50 years.

It is a fact that, during the last decade, the activity in conferences and expositions all over the world has reflected low activity in industrial manipulators and huge activity in other areas related with manipulation in unstructured environments and mobility, including wheeled, flying, underwater, legged and humanoid robots. Maybe the key is that new challenges in manipulation in factories require less research now because factory needs lie in the field of traditional engineering.

With these premises we can conclude: Yes, definitely robotics research is moving from industrial to field and service applications, and most robotics researchers are enthusiastic about this broad, exciting field. One development that is very representative of the way the field is evolving is the controversy set off by Prof. Engelberger, the creator of the first robotics company, at the 2005 International Robot Exhibition in Tokyo, Japan, when he commented on the needless research by both Japanese companies and scientific institutions for developing toy-like animal and humanoid robots for very doubtful use. Engelberger thus gained many detractors, who have rapidly argued back that these kinds of robots are a necessary step in the evolution towards real robots capable of helping disabled persons, performing dangerous work and moving in hazardous places.

Other defenders of the development of human-like personal robots advocate the importance of aiming at such challenging tasks because of the technology that can be developed, which would prove very important from the commercial point of view in other industrial activities.

Maybe behind all the arguments there still lies the human dream of the universal robot—a single device that can perform any task. Nothing better for that than a device resembling—what else?—a human being. So, let our imagination fly into the world of service robotics, but, please, do not forget to keep an eye on traditional industrial manipulators.

References

- [1] J.J. Craig, *Introduction to Robotics*. Reading, MA: Addison-Wesley, 2nd ed., 1989.
- [2] L. Zhenyu, C. Yinglin, and Q. Daokui, "Research on robot calibration," *Robot*, vol. 24, no. 5, pp. 447–450, 2002.
- [3] S. Lei, L. Jingtai, S. Weiwei, W. Shuihua, and H. Xingbo, "Geometry-based robot calibration method," in *Proc. IEEE Int. Conf. Robotics and Automation*, pp. 1907–1912, 2004.
- [4] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," *Int. J. Robot. Res.*, vol. 5, no. 1, pp. 90–98, 1986.
- [5] J.F. Canny, *The Complexity of Robot Motion Planning*, Cambridge, MA: MIT Press, 1988.
- [6] J.T. Schwartz and M. Sharir, "On the 'piano movers' problem: 1. The case of two-dimensional rigid polygonal body moving amidst polygonal barriers," *Commun. Pure Appl. Math.*, vol. 36, pp. 345–398, 1983.
- [7] J. Bobrow, S. Dubowsky, and J. Gibson, "Time-optimal control of robotic manipulators along specified paths," *Int. J. Robot. Res.*, vol. 4, no. 3, pp. 3–17, 1985.
- [8] J.J.E. Slotine and W. Li, *Applied Nonlinear Control*. Upper Saddle River, NJ: Prentice-Hall, 1991.
- [9] K.J. Aström and K.B. Wittenmark, *Adaptive Control*. Reading, MA: Addison-Wesley, 1989.
- [10] B. Siciliano and L. Villani, *Robot Force Control*. Norwell, MA: Kluwer, 1999.
- [11] P.C. Watson, "Remote center compliance system," U.S. Patent No. 4098001, Jul. 1978.
- [12] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2003.
- [13] R.S. Michalski, J.C. Carbonell, and T.M. Mitchell, *Machine Learning*, Palo Alto, CA: Tioga, 1983.
- [14] S.G. Tzafestas, "Fuzzy systems and fuzzy expert control: An overview," *Knowledge Eng. Rev.*, vol. 9, no. 3, pp. 229–268, 1994.
- [15] G. Niemeyer and J.J.E. Slotine, "Telemanipulation with time delays," *Int. J. Robot. Res.*, vol. 23, no. 9, pp. 873–890, 2004.
- [16] B. Davies, *A Discussion of Safety Issues for Medical Robots*, In R. Taylor, S. Lavalée, G. Burdea, and R. Moesges, Eds. *Computer-Integrated Surgery*. Cambridge, MA: MIT Press, pp. 287–296, 1996.
- [17] M. Blackwell, C. Nikou, A. DiGioia, and T. Kanade, "An image overlay system for medical data visualization," *Med. Image Anal.*, vol. 4, pp. 67–72, 2000.
- [18] R. Kumar, P. Berkelman, Gupta, A. Barnes, P.S. Jensen, L.L. Whitcomb, and R.H. Taylor, "Preliminary experiments in cooperative human/robot force control for robot assisted microsurgical manipulation," in *Proc. IEEE Int. Conf. Robotics Automation*, pp. 610–617, 2000, San Francisco, CA.
- [19] G.H. Ballantyne and F. Moll, "The da vinci telerobotic surgical system: The virtual operative field and telepresence surgery," *Surg. Clin. North Amer.*, vol. 86, no. 6, pp. 1293–1304, 2003.
- [20] J. Marescaux and F. Rubino, "The zeus robotic system: Experimental and clinical applications," *Surg. Clin. North Amer.*, vol. 86, no. 6, pp. 1305–1315, 2003.
- [21] Y. Kim and A.M. Cook, *Manipulation and Mobility Aids*, In J.G. Webster et al, Eds. *Electronic Devices for Rehabilitation*. London, U.K.: Chapman and Hall, 1985.
- [22] H.I. Krebs, B.T. Volpe, M.L. Aisen, and N. Hogan, "Increasing productivity and quality of care: Robot-aided neuro-rehabilitation," *J. Rehab. Res. Devel.*, vol. 37, no. 6, pp. 639–652, 2000.
- [23] K. Kiguchi and T. Fukuda, "A 3DOF exoskeleton for upper-limb motion assist—Consideration of the effect of bi-articular muscles," in *Proc. IEEE Int. Conf. Robotics Automation*, New Orleans, LA, pp. 2424–2429, 2004.
- [24] L.J. Leifer, "Rehabilitative robotics, the stanford robotic aid," in *Proc. WESCON*, San Francisco, CA, 1981, pp. 4–15.
- [25] G. Bolmsjö, H. Neveryd, and H. Efrting, "Robotics in rehabilitation," *IEEE Trans. Rehab. Eng.*, vol. 3, no. 1 pp. 77–83, Jan. 1995.
- [26] J. Leonard and H. Durrant-White, *Directed Sonar Sensing for Mobile Robot Navigation*. Norwell, MA: Kluwer, 1992.
- [27] R. Simmons and S. Koenig, "Probabilistic robot navigation in partially observable environments," in *Proc. Int. Joint Conf. Artificial Intelligence*, pp. 1080–1087, 1995.
- [28] W. Burgard, D. Fox, D. Hennig, and T. Schmidt, "Estimating the absolute position of a mobile robot using position probability grids," in *Proc. AAAI National Conf. Artificial Intelligence*, pp. 896–901, 1996.
- [29] S. Thrun, D. Fox, F. Dellaert, and W. Burgard, "Particle filters for mobile robot localization," A. Doucet, N. de Freitas and N. Gordon, eds. in *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, New York, 2001.
- [30] A. Elfes, "Sonar-based real-world mapping and navigation," *IEEE J. Robot. Automation*, vol. 3, no. 2 pp. 249–265, Mar. 1987.
- [31] M.J. Mataríé, "A distributed model for mobile robot environment-learning and navigation," M.S. thesis, Mass. Inst. Technol., Cambridge, MA, Jan. 1990, MIT AI Lab. Tech. Rep. AITR-1228, May 1990.
- [32] H.P. Moravec, "Sensor fusion in certainty grids for mobile robots," *AI Mag.*, vol. 9, no. 2, pp. 61–74, 1988.
- [33] R. Chatila and J.-P. Laumond, "Position referencing and consistent world modeling for mobile robots," in *Proc. IEEE Int. Conf. Robotics and Automation*, 1985, pp. 138–145.
- [34] J. Leonard, H.F. Durrant-White, and I.J. Cox, "Dynamic map building for an autonomous mobile robot," *Int. J. Robot. Res.*, vol. 11, no. 4, pp. 89–96, 1992.
- [35] J.A. Castellanos and J.D. Tardós, *Mobile Robot Localization and Map Building: A Multisensor Fusion Approach*, Norwell, MA: Kluwer, 2000.
- [36] J.E. Guivant and E. Nebot, "Optimization of the simultaneous localization and map-building algorithm for real-time implementation," *IEEE Trans. Robot. Automat.*, vol. 17, no. 3, pp. 242–257, May 2001.
- [37] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," in *Proc. AAAI National Conf. Artificial Intelligence*, Edmonton, Canada, 2002, AAAI, pp. 593–598.
- [38] F. Lu and E. Miliós, "Globally consistent range scan alignment for environment mapping," *Auton. Robots*, vol. 4, pp. 333–349, 1997.
- [39] Y. Liu and S. Thrun, "Results for outdoor-slam using sparse extended information filters," in *Proc. IEEE Int. Conf. Robotics Automation*, 2003, 1227–1233.
- [40] F. Dellaert, S.M. Seitz, C. Thorpe, and S. Thrun, "EM, MCMC, and chain flipping for structure from motion with unknown correspondence," *Mach. Learn.*, vol. 50, no. 1–2, pp. 45–71, 2003.
- [41] R. Biswas, B. Limketkai, S. Sanner and S. Thrun, "Towards object mapping in non-stationary environments with mobile robots," in *Proc. Int. Conf. Intelligent Robots Systems*, pp. 1014–1019, 2002.
- [42] D. Wolf and G.S. Sukhatme, "Online simultaneous localization and mapping in dynamic environments," in *Proc. IEEE Int. Conf. Robotics Automation*, 2004, pp. 1301–1306.
- [43] S. Thrun, "Robotic mapping: A survey," Tech. Rep. CMU-CS-02-111, School Comp. Sci., Carnegie Mellon Univ., Pittsburgh, PA 15213, Feb. 2002.
- [44] S. Thrun, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "MINERVA: A second generation mobile tour-guide robot," in *Proc. IEEE Int. Conf. Robotics Automation*, 1999, pp. 1999–2005.
- [45] J. Yuh, "Design and control of autonomous underwater robots: A survey," *Auton. Robots*, vol. 8, pp. 7–24, 2000.
- [46] D.A. Read, F.S. Hover, and M.S. Triantafyllou, "Forces on oscillating foils for propulsion and maneuvering," *J. Fluids Structures*, vol. 17, pp. 163–183, 2003.
- [47] J. Ayers, "Underwater walking," *Arthropod Struct. Devel.*, vol. 33, no. 3, 347–360, 2004.
- [48] G. Georgidas, A. German, A. Hogue, H. Liu, C. Prahacs, A. Ripsman, R. Sim, L.A. Torres, P. Zhang, M. Buehler, G. Dudek, M. Jenkin, and E. Miliós, "Aqua: An aquatic walking robot," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2004, Sendai, Japan, pp. 1731–1737.
- [49] R.B. McGhee and A.A. Frank, "On the stability properties of quadruped creeping gaits," *Math. Biosci.*, vol. 3, pp. 331–351, 1968.

- [50] P. Gonzalez de Santos, E. Garcia, and J. Estremera, *Quadrupedal Locomotion: An Introduction to the Control of Four-Legged Robots*, London, U.K.: Springer-Verlag, 2006.
- [51] S.M. Song and K.J. Waldron, *Machines that walk: The adaptive suspension vehicle*. Cambridge, MA: MIT Press, 1989.
- [52] T.A. McMahon, "The role of compliance in mammalian running gaits," *J. Experim. Biol.*, vol. 115, pp. 263–282, 1985.
- [53] M.H. Raibert, *Legged Robots That Balance*. Cambridge, MA: MIT Press, 1986.
- [54] K. Yoneda, H. Iiyama, and S. Hirose, "Intermittent trot gait of a quadruped walking machine dynamic stability control of an omnidirectional walk," in *Proc. IEEE Int. Conf. Robotics Automation*, 1996, pp. 3002–3007, Atlanta, GA.
- [55] H. Kimura and Y. Fukuoka, "Biologically inspired adaptive dynamic walking in outdoor environment using a self-contained quadruped robot: 'tekken2'," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2004.
- [56] S. Hirose and K. Kato, "Quadruped walking robot to perform mine detection and removal task," in *Proc. Int. Conf. Climbing Walking Robots*, 1998, pp. 261–266, Brussels, Belgium.
- [57] K. Nonami, Q.J. Huang, D. Komizo, N. Shimoi, and H. Uchida, "Humanitarian mine detection six-legged walking robot," in *Proc. Int. Conf. Climbing Walking Robots*, 2000, pp. 861–868, Madrid, Spain.
- [58] P. Gonzalez de Santos, E. Garcia, J. Estremera, and M.A. Armada, "DYLEMA: Using walking robots for landmine detection and location," *Int. J. Syst. Sci.*, vol. 36, no. 9, pp. 545–558, 2005.
- [59] P. Gonzalez de Santos, M.A. Armada, and M.A. Jimenez, "Ship building with ROWER," *IEEE Robot. Automat.Mag.*, vol. 7, no. 4, pp. 35–43, Dec. 2000.
- [60] R. Molfino, M. Armada, F. Cepolina, and M. Zoppi, "Roboclimber the 3 ton spider," *Ind. Robot: Int. J.*, vol. 32, no. 2, pp. 163–170, 2005.
- [61] Q. Huang, K. Yokoi, S. Kajit, K. Kaneko, H. Arai, N. Koyachi, and K. Tanie, "Planning walking patterns for a biped robot," *IEEE Trans. Robot. Automat.*, vol. 17, no. 3, pp. 280–289, May 2001.
- [62] C. Azevedo, P. Poignet, and B. Espiau, "Moving horizon control for biped robots without reference trajectory," in *Proc. IEEE Int. Conf. Robotics Automation*, 2002, pp. 2762–2767.
- [63] L. Roussel, C. Canudas de Wit, and A. Goswami, "Generation of energy optimal complete gait cycles for biped robots," in *Proc. IEEE Int. Conf. Robotics Automation*, 1998, pp. 2036–2041.
- [64] M. Vukobratovic, A.A. Frank, and D. Juricic, "On the stability of biped locomotion," *IEEE Trans. Biomed. Eng.*, vol. BME-17, no. 1, pp. 25–36, Jan. 1970.
- [65] J.H. Park, "Impedance control for biped robot locomotion," *IEEE Trans. Robot. Automat.*, vol. 17, no. 6, pp. 870–882, Nov. 2001.
- [66] I. de la Guia, P. Staroverov, M. Arbulu, and C. Balaguer, "Fast algorithm for kinematics problems solving of the low-cost legged robot LEROI," in *Proc. Int. Conf. Climbing Walking Robots*, 2002, pp. 775–782.
- [67] T. Takahashi and A. Kawamura, "Posture control for biped robot walk with foot toe and sole," in *Proc. IECON '01*, 2001, pp. 329–334.
- [68] Q. Huang, Y. Nakamura, and T. Inamura, "Humanoids walk with feedforward dynamic pattern and feedback sensory reflection," in *Proc. IEEE Int. Conf. Robotics Automation*, 2001, pp. 4220–4225.
- [69] P. Sardain and G. Bessonnet, "Forces acting on a biped robot. center of pressure zero moment point," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 34, no. 5, pp. 630–637, Oct. 2004.
- [70] K. Hirai, M. Hirose, Y. Haikawa, and T. Takenaka, "The development of Honda humanoid robot," in *Proc. IEEE Int. Conf. Robotics Automation*, May 1998, pp. 1321–1326, Leuven, Belgium.
- [71] D.G. Caldwell, N. Tsagarakis, P. Artrit, J. Canderle, S. Davis, and G.A. Medrano-Cerda, "Biomimetic and smart technology principles of humanoid design," in *Proc. IEEE/ASME Int. Conf. Advanced Intelligent Mechatronics*, 2001, pp. 965–970.
- [72] R. Caballero, T. Akinfiev, C. Manzano, H. Montes, and M. Armada, "Design of the SMART actuated ROBICAM biped robot," in *Proc. Int. Conf. Climbing Walking Robots*, 2002, pp. 409–416.
- [73] M. Gienger, K. Löffler, and F. Pfeiffer, "Towards the design of a biped jogging robot," in *Proc. IEEE Int. Conf. Robotics Automation*, 2001, pp. 4140–4145.
- [74] K. Nakadai and H. Tsujino, "Towards new human-humanoid communication: Listening during speaking by using ultrasonic directional speaker," in *Proc. IEEE Int. Conf. Robotics Automation*, 2005, pp. 1495–1500.
- [75] P. Nilas, P. Rani, and N. Sarkar, "An innovative high-level human-robot interaction for disabled persons," in *Proc. IEEE Int. Conf. Robotics Automation*, 2004, pp. 2309–2314.
- [76] Y. Chen and W.S. Newman, "A human-robot interface based on electrooculography," in *Proc. IEEE Int. Conf. Robotics Automation*, 2004, pp. 243–248.

Knowledge Discovery in Databases (KDD): An Overview

Nwagu, Chikezie Kenneth, Omankwu, Obinnaya Chinecherem, and Inyiama, Hycient

¹ Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,
Nwaguchikeziekeneth@hotmail.com

² Computer Science Department, Michael Okpara University of Agriculture, Umudike
Umuahia, Abia State, Nigeria
saintbeloved@yahoo.com

³ Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka
Anambra State, Nigeria.

ABSTRACT

Knowledge Discovery in Databases is the process of searching for hidden knowledge in the massive amounts of data that we are technically capable of generating and storing. Data, in its raw form, is simply a collection of elements, from which little knowledge can be gleaned. With the development of data discovery techniques the value of the data is significantly improved. A variety of methods are available to assist in extracting patterns that when interpreted provide valuable, possibly previously unknown, insight into the stored data. This information can be predictive or descriptive in nature. Data mining, the pattern extraction phase of KDD, can take on many forms, the choice dependent on the desired results. KDD is a multi-step process that facilitates the conversion of data to useful information. Our increased ability to gain information from stored data raises the ethical dilemma of how the information should be treated and safeguarded.

Keywords

Knowledge Discovery Databases, Data Mining, Knowledge Mining

1. INTRODUCTION

The desire and need for information has led to the development of systems and equipment that can generate and collect massive amounts of data. Many fields, especially those involved in decision making, are participants in the information acquisition game. Examples include: finance, banking, retail sales, manufacturing, monitoring and diagnosis, health care, marketing and science data acquisition. Advances in storage capacity and digital data gathering equipment such as scanners, has made it possible to generate massive datasets, sometimes called data warehouses that measure in terabytes. For example, NASA's Earth Observing System is expected to return data at rates of several gigabytes per hour by the end of the century (Way, 1991). Modern scanning equipment record millions of transactions from common daily activities such as supermarket or department store checkout-register sales. The explosion in the number of resources available on the World

Wide Web is another challenge for indexing and searching through a continually changing and growing "database."

Our ability to wade through the data and turn it into meaningful information is hampered by the size and complexity of the stored information base. In fact, the sheer size of the data makes human analysis untenable in many instances, negating the effort spent in collecting the data. There are several viable options currently being used to assist in weeding out usable information. The information retrieval process using these various tools is referred to as Knowledge Discovery in Databases (KDD).

"The basic task of KDD is to extract knowledge (or information) from lower level data (databases) (Fayyad *et al*, 1995). There are several formal definitions of KDD, all agree that the intent is to harvest information by recognizing patterns in raw data. Let us examine definition proposed by Fayyad, Piatetsky-Shapiro and Smyth, "Knowledge Discovery in Databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad *et al*, 1995). The goal is to distinguish from unprocessed data, something that may not be obvious but is valuable or enlightening in its discovery. Extraction of knowledge from raw data is accomplished by applying Data Mining methods. KDD has a much broader scope, of which data mining is one step in a multidimensional process.

Knowledge Discovery in Databases Process

Steps in the KDD process are depicted in the following diagram. It is important to note that KDD is not accomplished without human interaction. The selection of a data set and subset requires an understanding of the domain from which the data is to be extracted. For example, a database may contain customer address that would not be pertinent to discovering patterns in the selection of food items at a grocery store. Deleting non-related data elements from the dataset reduces the search space during the data mining phase of KDD. If the dataset can be analyzed using a sampling of the

data, the sample size and composition are determined during this stage.

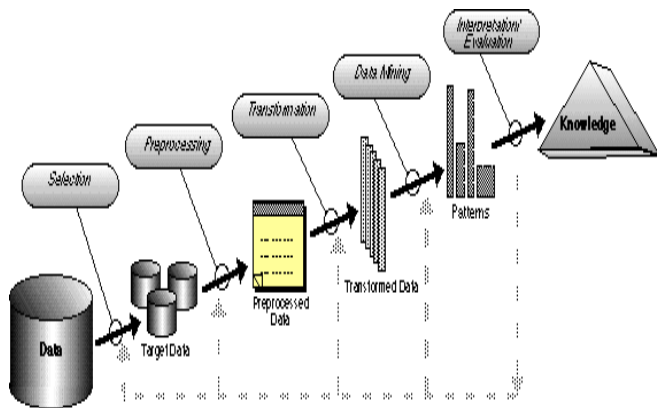


Fig. 1: Steps in KDD Process.

Databases are notoriously "noisy" or contain inaccurate or missing data. During the preprocessing stage the data is cleaned. This involves the removal of "outliers" if appropriate; deciding strategies for handling missing data fields; accounting for time sequence information, and applicable normalization of data (Fayyad, 1996)

In the transformation phase attempts to limit or reduce the number of data elements that are evaluated while maintaining the validity of the data. During this stage data is organized, converted from one type to another (i.e. changing nominal to numeric) and new or "derived" attributes are defined.

At this point the data is subjected to one or several data mining methods such as classification, regression, or clustering. The data mining component of KDD often involves repeated iterative application of particular data mining methods. "For example, to develop an accurate, symbolic classification model that predicts whether magazine subscribers will renew their subscriptions, a circulation manager might need to first use clustering to segment the subscriber database, and then apply rule induction to automatically create a classification for each desired cluster (Simoudis, 1996). Various data mining methods will be discussed in more detail in following sections.

The final step is the interpretation and documentation of the results from the previous steps. Actions at this stage could consist of returning to a previous step in the KDD process to further refine the acquired knowledge, or translating the knowledge into a form understandable to the user. A commonly used interpretive technique is visualization of the extracted patterns. The results should be critically reviewed and conflicts with previously believed or extracted knowledge resolved.

Understanding and committing to all phases of the data mining process is crucial to its success.

Data Mining Models

A few of the many model functions being incorporated in KDD include:

Classification: mapping or classifying data into one of several predefined classes (Hand, 1981). For example, a bank may establish classes based on debt to income ratio. The classification algorithm determines within which of the two classes an applicant falls and generates a loan decision based on the result.

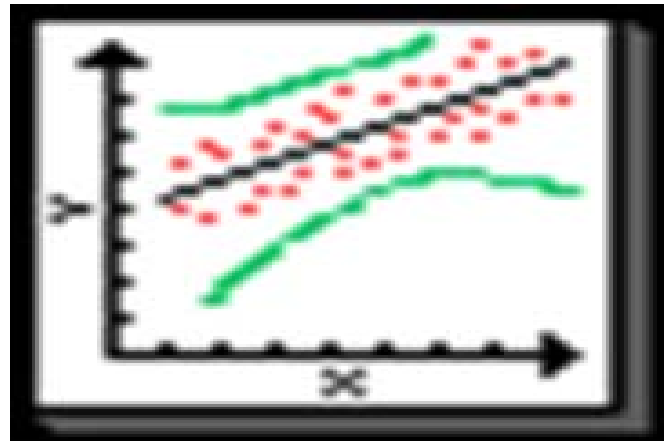


Figure 2: Regression Analysis

Regression: "a learning function which maps a data item to a real-valued prediction variable (Hand, 1981). Comparing a particular instance of an electric bill to a predetermined norm for that same time period and observing deviations from that norm is an example of regression analysis.

Clustering: "maps a data item into one of several categorical classes (or clusters) in which the classes must be determined from the data, unlike classification in which the classes are predefined. Clusters are defined by finding natural groupings of data items based on similarity metrics or probability density models (Fayyad *et al*, 1996). An example of this technique would be grouping patients based on symptoms exhibited. The clusters need not be mutually exclusive.

Summarization: generating a concise description of the data. Routine examples of these techniques include the mean and standard deviation of specific data elements within the dataset.

Dependency modeling: developing a model that shows how variables are interrelated. An example would be a model showing that electrical usage is highly correlated with the ambient temperature.

Choosing a Data Mining Model

There are no established guidelines to assist in choosing the correct algorithm to apply to a dataset. Typically, the more complex models may fit the data better but may also be more difficult to understand and to fit reliably (Fayyad *et al*, 1995). Successful applications often use simpler models due to their ease of translation. Each technique tends to lend itself to a particular type problem. Understanding the domain will assist in determining what kind of information is needed from the discovery process thereby narrowing the field of choice. Results can be broken into two general categories; prediction and description. Prediction, as the name infers, attempts to forecast the possible future values of data elements. Prediction is being applied extensively in the area of finance in an attempt to forecast movement in the stock market. Description seeks to discover interpretable patterns in the data. Fraud detection is an application that uses description to identify characteristics of potential fraudulent transactions.

Classification, clustering, summarization and dependency modeling are descriptive models, while regression is predictive.

Current Applications of KDD

Several Knowledge Discovery Applications have been successfully implemented. "SKICAT, a system which automatically detects and classifies sky objects image data resulting from a major astronomical sky survey. SKICAT can outperform astronomers in accurately classifying faint sky objects(Fayyad *et al*, 1995). KDD is being used to flag suspicious activities on two frontiers: Falcon alerts banks of possible fraudulent credit card transactions and the FAIS system being employed by the Financial Crimes Enforcement Network detects financial transactions that may indicate money laundering (Simoudis, 1996). Market Basket Analysis (MBA) has incorporated discovery driven data mining techniques to gain insights about customer behavior. Other applications are being used in the Molecular Biology, Global Climate Change Modeling and other concentrations where the volume of data exceeds our ability to decipher its meaning.

Privacy Concerns and Knowledge Discovery

Although not unique to Knowledge Discovery, sensitive information is being collected and stored in these huge data warehouses. Concerns have been raised about what information should be protected from KDD-type access. The ethical and moral issues of invasion of privacy are intrinsically connected to pattern recognition. Safeguards are being discussed to prevent misuses of the technology.

Summary

Knowledge Discovery in Databases is answering a need to make use of the mountains of data that is accumulating daily. KDD enlists the power of computers to assist in the

recognizing patterns in data, a task that exceeds human ability as the size of data warehouses increase. New methods of analysis and pattern extraction are being developed and adapted to KDD. Which method is used depends on the domain and results expected. The accuracy of the recorded data must not be overlooked during the KDD process. Domain specific knowledge assists with the subjective analysis of KDD results. Much attention has been given to the data mining phase of KDD but earlier steps, such as data cleaning, play a significant role in the validity of the results.

The potential benefits of discovery driven data mining techniques in extracting valuable information from large complex databases are unlimited. Successful applications are surfacing in industries and areas where data retrieval is outpacing man's ability to effectively analyze its content. Users must be aware of the potential moral conflicts to using sensitive information.

REFERENCES

- (1) Way, J.; and Smith, E.A. "The evolution of Synthetic Radar Systems and Their Progression to the EOS SAR." IEEE Trans. Geoscience and Remote Sensing. Vol 29. No. 6. 1991. Pp962-985.
- (2) Fayyad, U.; Simoudis, E.; "Knowledge Discovery and Data Mining Tutorial MA1" from Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95) July 27, 1995 www-aig.jpl.nasa.gov/public/kdd95/tutorials/IJCAI95-tutorial.html
- (3) Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P; "From Data Mining to Knowledge Discovery: An overview" in Advances in Knowledge discovery and Data Mining. Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P; Uthurusamy, R. MIT Press. Cambridge, Mass.. 1996 pp. 1-36
- (4) Fayyad, U. "Data Mining and Knowledge Discovery: Making Sense Out of Data" in IEEE Expert October 1996 pp. 20-25
- (5) Simoudis, E. "Reality Check for Data Mining" in IEEE Expert October 1996 pp. 26-33
- (6) Hand, D. J. 1981 Discrimination and Classification. Chichester, U.K.: John Wiley and Sons
- (7) Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P; "From Data Mining to Knowledge Discovery: An overview" in Advances in Knowledge discovery and Data Mining. Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P; Uthurusamy, R. MIT Press. Cambridge, Mass.. 1996 pp. 1-36
- (8) Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P.; "The KDD Process for Extracting Useful Knowledge from Volumes of Data" in Communications of the ACM, November 1996/Vol 39, No.11 pp.27-34

(9) Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P; "From Data Mining to Knowledge Discovery: An overview" in Advances in Knowledge discovery and Data Mining. Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P; Uthurusamy, R. MIT Press. Cambridge, Mass.. 1996 pp. 1-36

(10) Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P; "From Data Mining to Knowledge Discovery: An overview" in Advances in Knowledge discovery and Data Mining. Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P; Uthurusamy, R. MIT Press. Cambridge, Mass.. 1996 pp. 1-36

(11) Simoudis, E. "Reality Check for Data Mining" in IEEE Expert October 1996 pp. 26-33

Knowledge Management (KM): An Overview

Omankwu, Obinnaya Chinecherem; Nwagu, Chikezie Kenneth, and Inyiama, Hycient

¹ Computer Science Department, Michael Okpara University of Agriculture, Umudike
Umuahia, Abia State, Nigeria
saintbeloved@yahoo.com

² Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,
Nwaguchikeziekeneth@hotmail.com

³ Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka
Anambra State, Nigeria.

ABSTRACT

While knowledge management (KM) is becoming an established discipline with many applications and techniques, its adoption in health care has been challenging. Though, the health care sector relies heavily on knowledge and evidence based medicine is expected to be implemented in daily health care activities; besides, delivery of care relies on cooperation of several partners that need to exchange their knowledge in order to provide quality of care. This publication will give an overview of KM, its methods and techniques.

Keywords

Knowledge Management, Data Mining, Knowledge Mining

1. INTRODUCTION

In service base companies, knowledge is a central intangible asset; knowledge management deals with the creation, use, reuse, dissemination of Knowledge. Knowledge Management (KM) became a discipline during the 80's, and the growing role of information technologies enabled the development of efficient KM tools using databases and collaborative software.

KNOWLEDGE MANAGEMENT A

Brief History

Knowledge management had always been a central question in human societies. Indeed, its roots are to be found in the early history of human societies. Philosophers, Western as well as Eastern, have focused their attention on the question of

knowledge; already in ancient Greece, 'scientific' discussions often lead to philosophical debates, especially on the concept of knowledge. The creation of epistemology has finally formalized the question of knowledge; indeed, epistemology addresses primarily the question of "what is knowledge?" and discusses its creation and adoption. In the current discipline of knowledge management, philosophical considerations from several schools are taken into account, especially in the ontological knowledge management field (Grenon, 2003).

On the other hand, practical knowledge management has always taken place in the society, and transmission of knowledge was much related to the technical progress. Beginning in the middle age, knowledge transmission occurred under what was called "Wandergesellen" in Germany and "Compagnonnage" in France, where craftsmen and artisan take a tour of the country for 6 months or one year to learn from several masters. This was one of the first structured methodologies for tacit knowledge transmission. Knowledge first spread orally, then in writing; but it was restricted to a low circle of educated people till the development of printing. If the first printing focused on religious and literature purpose, technical and specialized books began to spread after the wide adoption of the printing press.

In the 20th century, management as well as cognitive sciences and psychology led to today's Knowledge Management (KM) (Wiig, 1999). The current situation of KM started in the 1980s with the wide use of information technologies in companies; the focus was on the intangible asset that knowledge represents. The word KM itself appeared in the 80s and the academic discipline was created in 1995 (Stankosky, 2005).

Goals and challenges of KM are many; for companies, KM should increase their performance, help to develop partnerships, evaluate risks, organize management and enhance their economic value. Development of corporate memory and measurement tools also aims at assessing intangible assets in the companies. Besides, knowledge transfer enhancement and companies' performance evaluations became issues of major importance. After twenty productive years in KM, the first criticisms appeared in 2002. T.D. Wilson (Wilson, 2002) discussed the foundation of KM, mainly because of the difficulty to distinguish information from knowledge in most KM theories. He drew the conclusion that KM was a management fad and should disappear in the upcoming years. Actually KM survived well those criticisms, even if the precision of the vocabulary is not comparable to the one used in epistemology or in computer science based KM; probably the reason lies in the real need for companies as well as public organizations to use KM methods.

We can distinguish 2 main KM trends: people and information management centered and information technology centered. We should also recognize two other main orientations, the first is the need of evaluation in terms of performance measurement, and the second is the measurement of knowledge assets in order to evaluate the value of an organization (Wiig, 1999).

KM Frameworks

Frameworks for KM support are based on considerations related to the structure of knowledge and to the structure of organizations where the frameworks are applied. In most of models, knowledge types are determined based on different criteria, such as having structured or unstructured knowledge, and having tacit or explicit knowledge.

First we have to make a distinction between high level frameworks and implementation oriented ones. The latter one focus on the "how to" implement KM in an organization, whereas the first one discuss the question of "what is KM" (Wong & Aspinwall, 2004). As our purpose is to focus more on the "how to" question, we will focus in the next paragraphs on the implementation oriented frameworks.

High level frameworks discuss how to fill the gap between theory and practice, that is the case of Knowledge Creation Frameworks for example (Siebert, 2005).

Nonaka and Takeuchi (Nonaka & Takeuchi, 1995) depict steps to create knowledge in practice that go from perception to representation and from tacit

knowledge to explicit one; they also show how those steps can enhance company's efficiency.

Concerning implementation frameworks, Sunassee and Sewry (Sunassee & Sewry, 2002) defined three categories of frameworks: prescriptive, descriptive and hybrids. *Prescriptive* frameworks give direction concerning the procedures that should be used, without describing precisely their content or implementations, for example step approach frameworks are mainly prescriptive frameworks (Wong & Aspinwall, 2004). *Descriptive* frameworks describe the key factors of KM that can drive KM initiatives to success or to failure, their forms of representations are mostly graphical (Wong & Aspinwall, 2004); examples of descriptive frameworks can be found in (Gore 1. Gore, 1999; Holsapple & Joshi, 2002). Finally Hybrid approaches combine both prescriptive and descriptive methods.

It is important to find a way to compare KM frameworks; though, frameworks are dedicated to specific applications which make their comparison complicated. Wong and Aspinwall (Wong & Aspinwall, 2004) proposed a comparison method of frameworks based on four elements, their structure, the knowledge types they represent, the KM processes and the KM influences or factors.

Methods and Techniques in KM

We can categorize the methods and techniques in KM in three groups: people and technology, requirements elicitation and value measurement.

People and Technology

Early approaches of KM frameworks in the early 1990s mainly focused on the structural organization and IT solutions to improve knowledge management (Wiig, 1999). Those methods were adapted for slow moving businesses where goals and technical solutions are perfectly identified and the market does not evolve quickly. But, these approaches were not adapted in a subsequent fast moving business environment where new challenges started arise as fast as they disappear.

Human centered KM has been early identified and became a new school of thought, in the early 1990s. Peters (Peters, 1994) wrote "the answer turns out to lie more with psychology and marketing of knowledge within the family than with bits and bytes". Nowadays frameworks take both *human* and *technical* perspectives into account. We will discuss both approaches separately and show how both are integrated in nowadays frameworks. <https://sites.google.com/site/ijcsis/>
ISSN 1947-5500

Human Perspective: Motivation and Adoption

The main issue for companies is to motivate employees to use KM systems. Not only that the technology matters, but people implication in KM initiatives is a key factor for its success. Without incentives, employees are not ready to share their knowledge; therefore, numerous solutions have been proposed to motivate employees to make use of KM systems. Some companies provide financial incentives (monetary rewards) or non-financial incentives (air miles, days off) for the first users of the KM system. Incentives, financial or not, are particularly efficient in organization where detaining knowledge is often considered as a source of power. In addition to individual incentives, Zand (Zand, 1997) suggests a collaborative win-win reward system, in which a gain for an individual can be a gain for his peers, in opposition with classical win-lose rewards system.

It has also been recognized that higher management should use the system too; Liebowitz (Liebowitz, 1999) cites the success of the KM network of Buckman Labs, which was mostly a result of the high level implication of the senior management and especially the CEO.

The second motivation related issue is *knowledge adoption*; it has been a challenge that people were not ready to use or apply knowledge developed by others. Sussman and Siegal (Sussman & Siegal, 2003) built a theoretical model to understand the underlying issues of knowledge adoption; their study discussed the role of informational influence in the process of knowledge adoption, and showed the importance of the source *credibility* to convince people of the usefulness of the acquired knowledge. Once again, the commitment of senior management, who are trusted in their organization, can have a huge influence on the success of a project.

Technical Perspective: Data Mining, Inference Engine and Multi-Agent Systems

KM tools deals with explicit knowledge, meaning that Knowledge can be written on a support that is mainly an electronic one. Historically, collaborative tools, such as Lotus Notes, have been developed in the 1990s to enhance KM. Recent corporate tools widely adopted Web 2.0 technologies such as wiki platforms, semantic widgets, tagging and so on. Several concepts from the broad computer science research, such as *data mining*, *rules based reasoning*, and *multi-agent systems* have been integrated in KM solutions, the integration of those tools depends on the processes in action.

For instance, computer assisted Knowledge Discovery is mainly based on data mining techniques. A brief look on the papers of the Knowledge Discovery and Data mining (KDD) conference (Li, Liu, & Sarawagi, 2008) - the major conference on Knowledge Discovery - gives an overview of the overwhelming presence of data mining within Knowledge Discovery.

On the other hand, knowledge representation uses ontological models; due to the development of powerful inferences engines. Those representations can be used to infer knowledge from existing one, and shore up Knowledge Discovery processes. Several KM frameworks are based on ontologies (Fensel, 2002; Stojanovic, 2003; Sure, 2002; S.-Y. Yang, Lin, Lin, Cheng, Soo, 2005), since high level representation of Knowledge using ontologies enables powerful queries and Knowledge manipulation, retrieval and discovery.

Finally, the multi-agents system (MAS) paradigm is rightly suited to model the distribution of knowledge on autonomous entities, thus, it is used in order to disseminate knowledge among employees in organizations. MASs also take in account reactivity (adaptation to changes in an environment) and proactivity (anticipation of user needs and consequently taking initiatives). These two factors are the keys for the success of a KM project; indeed, KM initiatives require adapting quickly and being able to handle user needs. In this context, Virtual Knowledge Communities (Maret, Subercaze, & Calmet, 2008) present an efficient way to model KM in organization since it integrates the MAS approach and the ontological representation of Knowledge. Virtual model Knowledge Communities' model has been used for business (Subercaze, Pawar, Maret, & Calmet, 2008) as well as for health care purposes (El Morr, Subercaze, Maret, & Rioux, 2008).

Requirements Elicitation

Requirements can be seen under two angles, a technological one and a human centered one.

From the technological stand point, *storage* of Electronic Knowledge Repository represented a challenge at the early stages of KM; indeed hardware investment can require significant amount of money for a huge amount of data to process. Knowledge Discovery processes also require high computational power; nevertheless with the reduction of hardware costs, storage is no more a critical issue, but the latest research using ontological representation, inference engines, and data mining techniques still required powerful *computational power* (Guo, Pan, & Heflin, 2005).

Value Measurement

Assessing the value of KM is a primary concern for organizations. Like other intangibles assets, the reliability of Knowledge Management measurement in an organization is subject to debate. As underlined in a study for the European Union (Zambon, 2003), internal evaluations based on information provided by managers may be subject to bias and tend to overestimate the value of KM. On the other hand, evaluations conducted by third parties may be imprecise, as third parties may not have access to the internal knowledge assets. The absence of a market for intangible asset can also be a root of evaluation bias; indeed, knowledge as an intangible asset will be evaluated and appear on the financial report but cannot be sold and has no proper market value. Therefore, there is no market structure that can regulate knowledge evaluation. Several methods have been developed to estimate the value of knowledge in an organization, Skandia is the first company to have dealt with the *Intellectual Capital* (IC) measurement (N. Bontis, 1996). It defined Intellectual Capital as the sum of the human and structural Capital. Human capital combines abilities, knowledge, and innovation potential of the company's employees; it includes the company's philosophy and culture too. This kind of capital is not property of the company, but the company drives benefits out of it. Structural capital is the patents, trademarks, hardware and "everything that gets left behind when employees go home" (Nick Bontis, 2001). IC reports developed by Skandia used 36 metrics to give a monetary value to an organization; metrics includes customer satisfaction, satisfied employees, number of patents, annual turnover. Second generation methods such as *IC-index*, was an extension of the Skandia IC metric, it tried to merge the different indicators of Skandia into a single index (Roos, Roos, Edvinsson, & Dragonetti, 1997). Other metrics were developed to evaluate Knowledge Management Systems (KMS), Kankanhalli et al. (Kankanhalli & Tan, 2004) present a thorough review of KMS metrics.

REFERENCES

Abidi, S. S. R. (2001). Knowledge management in healthcare: towards 'knowledge-driven' decision-support services. *International Journal of Medical Informatics*, 63(1-2), 5–18.

Ahmad, R., Kausar, A. R., & David, P. (Writer) (2007). *The social management of embodied knowledge in a knowledge community*.

Anderson, R. A., & McDaniel, R. R. (2000). Managing health care organizations: where professionalism meets complexity science. *Health Care Management Review*, 25(1), 83–92.

Andreas, R. (2005). Three-dozen knowledge-sharing barriers managers must consider. *Journal of Knowledge Management*, 9(3), 18–35.

Andreas, R., & Nicholas, L. (2006). Knowledge management in the public sector: stakeholder partnerships in the public policy development. *Journal of Knowledge Management*, 10(3), 24.

Ansell, C. (2007). Fostering Innovation and Collaboration. *Medical Device Technology*, 18(1), 52.

Bali, R. K., & Dwivedi, A. N. (Eds.). (2007). *Healthcare Knowledge Management*: Springer.

Batalden, P., & Splaine, M. (2002). What will it take to lead the continual improvement and innovation of health care in the twenty-first century? *Quality Management in Health Care*, 11(1), 45.

Bate, S. P., & Robert, G. (2002). Knowledge management and communities of practice in the private sector: lessons for modernizing the National Health Service in England and Wales. *Public Administration*, 80(4), 643–663.

Bates, D. W., Spell, N., Cullen, D. J., Burdick, E., Laird, N., & Petersen, L. A. (1997). The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. *Journal of the American Medical Association*, 277(4), 307–311. doi:10.1001/jama.277.4.307

Bontis, N. (1996). There's a price on your head: Managing Intellectual Capital Strategically. *Business Quarterly*, (Summer): 40–47.

Bontis, N. (2001). Assessing knowledge assets: a review of models used to measure intellectual capital. *International Journal of Management Reviews*, 3(1), 41–60.

Brian, D. P. (2006). *Final Report of the Federal Advisor on Wait Times*. Ottawa, Canada: Health Canada.

- Buchan, I. E., & Hanka, R. (1997). Exchanging clinical knowledge via Internet. *International Journal of Medical Informatics*, 47(1-2), 39–41.
- Burnett, S. M., Williams, D. A., & Webster, L. (2005). Knowledge support for interdisciplinary models of healthcare delivery: a study of knowledge needs and roles in managed clinical networks. *Health Informatics Journal*, 11(2), 146–160.
- Caldwell, D. F., Chatman, J., O'Reilly, C. A., III, Ormiston, M., & Lapid, M. (2008). Implementing strategic change in a health care system: the importance of leadership and change readiness. *Health care management review*, 33(2), 124(110).
- Canadian Health Services Research Foundation. (2003). *The theory and practice of knowledge brokering in Canada's health system*. Ottawa, Canada: Canadian Health Services Research Foundation.
- Canadian Institute for Health Information. (2007). *Health Care in Canada*. Ottawa, Ontario: Canadian Institute for Health Information.
- Canongia, C., Antunes, A., de Nazare, M., & Pereira, F. (2004). Technological foresight - the *cinc.in* (2008). Retrieved October 10, 2008, from <http://cyn.in/>
- Cullen, D. J., Sweitzer, B. J., Bates, D. W., Burdick, E., Edmondson, A., & Leape, L. L. (1997). Preventable adverse drug events in hospitalized patients: A comparative study of intensive care and general care units. *Critical Care Medicine*, 25(8), 1289–1297.

PERSPECTIVES FOR THE USE OF KM IN HEALTH CARE.

Omankwu, Obinnaya Chinecherem; Nwagu, Chikezie Kenneth, and Inyiama, Hycient

¹ **Computer Science Department, Michael Okpara University of Agriculture, Umudike
Umuahia, Abia State, Nigeria
*saintbeloved@yahoo.com***

² **Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,
*Nwaguchikeziekeneth@hotmail.com***

³ **Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka
Anambra State, Nigeria.**

ABSTRACT

Knowledge management (KM) is becoming an established discipline with many applications and techniques, its perspective view in health care has been challenging. Though, the health care sector relies heavily on knowledge and evidence based medicine is expected to be implemented in daily health care activities; besides, delivery of care relies on cooperation of several partners that need to exchange their knowledge in order to provide quality of care. This publication will give a perspective view of KM in Health care.

Keywords

Knowledge Management, Data Mining, Knowledge Mining

1. INTRODUCTION

In service base companies, knowledge is a central intangible asset; knowledge management deals with the creation, use, reuse, dissemination of Knowledge. Knowledge Management (KM) became a discipline during the 80's, and the growing role of information technologies enabled the development of efficient KM tools using databases and collaborative software.

Beside the current knowledge management roles in the health care sector, few perspectives present an opportunity to develop new health care KM applications. These perspectives are virtual communities, mobility, Electronic Health Record (E.H.R.), and public health.

Virtual Communities

“Virtual” health care providers of different disciplines (e.g. medicine, nursing, social work, physical therapy, etc.) can create teams in which they combine their knowledge and expertise to provide a comprehensive plan of care. Though, it is essential to include patients in virtual health care teams; indeed, patients must be well informed about their conditions, treatment options and how to access them and be actively involved in their treatment (Davis, Wagner, & Groves, 2000). Health Virtual Communities, that include care givers and patients, in order to create manage and coordinate virtual medical teams (Pitsillides, et al., 2004).

Once a Health VC is in place, new knowledge emerges through social interactions (Ahmad, Kausar, & David, 2007). Patients have tacit knowledge about their medical condition and the way they experience their conditions, this tacit knowledge constitute a mine of information for clinical practice; indeed, it allows to get insight into the patient experience and hence assess her/ his quality of life as well as the impact of a drug on a person's life. Virtual communities in this respect constitute an opportunity for a holistic approach to clinical practice.

Besides, Health VCs constitute an opportunity for e-continuing education. In health care, continuous education is essential; some professionals cannot continue practising unless they undergo a yearly continuous education course in order to update their knowledge. In this context, knowledge based Health VCs can play a major role by pro-viding a platform for e-education and knowledge exchange between peers. The creation of

virtual network of experts opens the road to test new kind of cooperation paradigms and *peer-to-peer e-educational* paradigms (van Dijk, Hugenholtz, A-Tjak, & Schreinemakers, 2006).

Mobility

While managing knowledge will become an important daily practice, the future will be more mobile. We're witnessing already the explosion of mobile interactive devices, mobile health facilities, and the proliferation of e-homecare solutions (Hubert, 2006). Mobile knowledge management is the next step in mobile health care situations where patient is away from the point of care (O'Sullivan, McLoughlin, Bertolotto, & Wilson, 2007). The mobility approach is extremely valid in virtual communities (El Morr, 2007; Christo El Morr & Jalal Kawash, 2007; C. El Morr & J. Kawash, 2007); consequently, the creation of mobile Health VCs where knowledge is generated, disseminated and shared by both patients and caregivers is a next step that can provide advantage for both patients and caregivers (Hubert, 2006; Moreno & Isern, 2002; Siau & Shen, 2006).

Electronic Health Record (E.H.R.)

Worldwide, governments are striving to build national wide E.H.R. systems. There has been progress in this direction mainly in developed countries. Once health records are computerized the need will be to reach the right information about a patient at the right time, and to use the E.H.R. data for diagnosis purposes, for personal health decision support, for public health decision support, and for research purposes as well. Though, much of what has been done till now in E.H.R. involves data processing mainly (Van Vlymen, De Lusignan, Hague, Chan, & Dzregah, 2005); besides, health service managers are facing many difficulties when trying to access relevant data routinely for quality improvement (De Lusignan, Wells, Shaw, Rowlands, & Crilly, 2005). KM techniques can play here two roles one for managers and one for practitioners; indeed, KM techniques can help in searching for knowledge

(2) in the mass of data gathered helping practitioners to find more effective ways to treat patients by searching for similar patient case histories (O'Sullivan, et al., 2007), and helping managers to get relevant knowledge for total quality management (TQM) (McAdam Leonard, 2001). Establishing, electronic health records, per se, constitute only a first step; using the mass of data gathered in order to support practitioners in generating knowledge and providing quality practise is the challenge ahead.

Evidence-Based Public Health

Networks for health care surveillance continue to evolve (Health Canada, 1999); nevertheless, studies show that information and communication technology are less used in public health than in other sectors of the society (Goddard, et al., 2004; Revere, et al., 2007). Public health is traditionally data processing and data analysis oriented, though there is more awareness that a shift is needed in public health from data driven decision making to knowledge driven decision making, or to put it in Goddard et al. words "pro-vide direct guidance on the relative effectiveness of different interventions in a specific situation" (Goddard, et al., 2004). KM can play a vital role in organizing, structuring and supporting *evidence based public health* decision making (Andreas & Nicholas, 2006; Revere & Fuller, 2008). In this context, research needs to unveil how the public health community communicates and cooperate, particularly in terms of role and communication strategies, artifact used, etc. Different profiles of knowledge health care workers can then be sketched. Research methods from the Computer Supported Collaborative Work (CSCW) field can be used. Findings can well be integrated in the context of Community of Practice where knowledge tools can further knowledge creation, communication and transfer. The medical field is experiencing a move to evidence based medicine, a similar move to *evidence-based public health* is important and would be strategic for an advanced management of population health; KM can play a vital role to make this move.

Knowledge Transfer

Knowledge transfer is concerned with dissemination of knowledge connecting and adapting research findings to the society needs. Increasingly, the role of knowledge broker is recognized as vital in knowledge transfer (Lind & Persborn, 2000); knowledge brokering “links decision makers with researchers, facilitating their interaction” (Canadian Health Services Research Foundation, 2003; Lomas, 2007). In this context, there is a crucial need to understand how knowledge is transferred, and transformed while it is transferred (Syed, 1999); cognitive theory can be of much help in this domain. This understanding will help providing a feedback to knowledge generators (i.e. researchers) and widen their knowledge (i.e. help generating more knowledge) (Figure 1).

Health 2.0 and Semantic Web

The term health 2.0 embeds the concepts of healthcare, web 2.0 and e-health. Following the web 2.0 principles, health 2.0 is driven by participatory ideas. In health 2.0, each actor of the system, patients, stakeholders are involved in the process of amelioration of the health care system using existing web 2.0 social networking, semantic web and collaborative tools.

As well as web 2.0, health 2.0 is an imprecise term. Most of the applications are focused on enhancing communication in the community. For instance, Sermo1 is physicians community dedicated for information exchange and collaboration’, and DoubleCheckMD2 is a patient oriented applications dedicated for drugs side effects; while Vitals3 help patients find a relevant doctor matching search criteria and write reviews on doctors, and PatientsLikeMe4 is an online community for patient with life threatening conditions. In health 2.0, allow patients to share their experience, medical data with other patients, doctors and research organizations; it aims at establishing data-sharing partnerships.

The trend in health 2.0 is to enhance collaborations, either between physicians or between patients, and to create new relationships between patients and doctors and research organizations.

On the other hand, semantic Web technologies enable a next step in accessing data at the scale of the web; indeed, RDF and OWL technologies are being used for knowledge modeling and for large database integrations. Currently, the W3C Semantic Web in Health Care and Life Sciences Interest Group (HCLSIG) aims at offering a better access.

to information from many domains and processes for an efficient decision support and disease management. Initiatives like OBI (Ontology for Biomedical Investigations) or RNA Ontology Consortium are the results of the movement initiated by the HCLSIG.

While, the current health 2.0 applications are based on relational databases; we believe that in the near future, we will see a merger between health 2.0 and the semantic web technologies developed by HCLSIG. The resulting applications could fairly improve automated knowledge management related to healthcare.

CONCLUSION

Knowledge management in health care is progressing; the complexity and challenges facing the health care sector can be addressed by adopting of KM strategies.

The use of KM in health care is promising to enhance the quality of care for patients by providing them with a continuity of care. The implementation of Health care KM system will allow health care partners (e.g. practitioners, administrators, etc.) to conduct evidence based practice and to collaborate relying on the best knowledge available.

This is a challenge that opens the way to more innovations in both KM and health. The current state of KM in health care can be improved; we believe that new practices such as, health 2.0 applications, VCs and evidence based medicine will help to increase the global quality of care of the patients as well as the efficiency of KM in healthcare.

REFERENCES

- Abidi, S. S. R. (2001). Knowledge management in healthcare: towards 'knowledge-driven' decision-support services. *International Journal of Medical Informatics*, 63(1-2), 5–18.
- Ahmad, R., Kausar, A. R., & David, P. (Writer) (2007). *The social management of embodied knowledge in a knowledge community*.
- Anderson, R. A., & McDaniel, R. R. (2000). Managing health care organizations: where professionalism meets complexity science. *Health Care Management Review*, 25(1), 83–92.
- Andreas, R. (2005). Three-dozen knowledge-sharing barriers managers must consider. *Journal of Knowledge Management*, 9(3), 18–35.
- Andreas, R., & Nicholas, L. (2006). Knowledge management in the public sector: stakeholder partnerships in the public policy development. *Journal of Knowledge Management*, 10(3), 24.
- Ansell, C. (2007). Fostering Innovation and Collaboration. *Medical Device Technology*, 18(1), 52.
- Bali, R. K., & Dwivedi, A. N. (Eds.). (2007). *Healthcare Knowledge Management*: Springer.
- Batalden, P., & Splaine, M. (2002). What will it take to lead the continual improvement and innovation of health care in the twenty-first century? *Quality Management in Health Care*, 11(1), 45.
- Bate, S. P., & Robert, G. (2002). Knowledge management and communities of practice in the private sector: lessons for modernizing the National Health Service in England and Wales. *Public Administration*, 80(4), 643–663.
- Bates, D. W., Spell, N., Cullen, D. J., Burdick, E., Laird, N., & Petersen, L. A. (1997). The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. *Journal of the American Medical Association*, 277(4), 307–311. doi:10.1001/jama.277.4.307
- Bontis, N. (1996). There's a price on your head: Managing Intellectual Capital Strategically. *Business Quarterly*, (Summer): 40–47.
- Bontis, N. (2001). Assessing knowledge assets: a review of models used to measure intellectual capital. *International Journal of Management Reviews*, 3(1), 41–60.
- Brian, D. P. (2006). *Final Report of the Federal Advisor on Wait Times*. Ottawa, Canada: Health Canada.
- Buchan, I. E., & Hanka, R. (1997). Exchanging clinical knowledge via Internet. *International Journal of Medical Informatics*, 47(1-2), 39–41.
- Burnett, S. M., Williams, D. A., & Webster, L. (2005). Knowledge support for interdisciplinary models of healthcare delivery: a study of knowledge needs and roles in managed clinical networks. *Health Informatics Journal*, 11(2), 146–160.
- Caldwell, D. F., Chatman, J., O'Reilly, C. A., III, Ormiston, M., & Lapiz, M. (2008). Implementing strategic change in a health care system: the importance of leadership and change readiness. *Health care management review*, 33(2), 124(110).
- Canadian Health Services Research Foundation. (2003). *The theory and practice of knowledge brokering in Canada's health system*. Ottawa, Canada: Canadian Health Services Research Foundation.
- Canadian Institute for Health Information. (2007). *Health Care in Canada*. Ottawa, Ontario: Canadian Institute for Health Information.
- Canongia, C., Antunes, A., de Nazare, M., & Pereira, F. (2004). Technological foresight - the *cinc.in* (2008). Retrieved October 10, 2008, from <http://cyn.in/>
- Cullen, D. J., Sweitzer, B. J., Bates, D. W., Burdick, E., Edmondson, A., & Leape, L. L. (1997). Preventable adverse drug events in hospitalized patients: A comparative study of intensive care and general care units. *Critical Care Medicine*, 25(8), 1289–1297.

Robotics and Artificial Intelligence: Differences and Similarities

Omankwu, Obinnaya Chinecherem, Nwagu, Chikezie Kenneth, and Inyama, Hycient

¹ Computer Science Department, Michael Okpara University of Agriculture, Umudike
Umuahia, Abia State, Nigeria
saintbeloved@yahoo.com

² Computer Science Department, Nnamdi Azikiwe University, Awka Anambra State, Nigeria,
Nwaguchikeziekeneth@hotmail.com

³ Electronics & Computer Engineering Department, Nnamdi Azikiwe University, Awka
Anambra State, Nigeria.

ABSTRACT

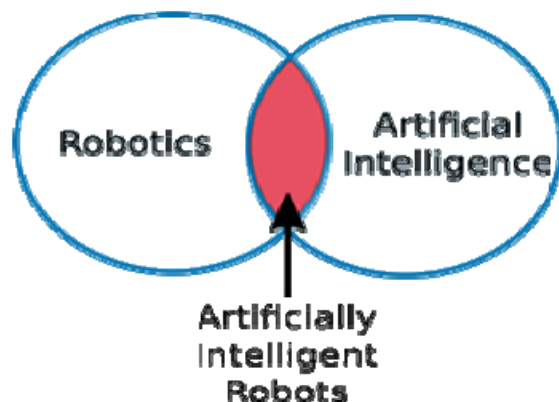
Robotics and artificial intelligence serve very different purposes. However, people often get them mixed up. Robotics is a branch of technology which deals with robots. Robots are programmable machines which are usually able to carry out a series of actions autonomously, or semi-autonomously. Artificial intelligence (AI) is a branch of computer science. It involves developing computer programs to complete tasks which would otherwise require human intelligence.

Keywords: Robotics, artificial intelligence, algorithm.

Introduction

Robotics and artificial intelligence serve very different purposes. However, people often get them mixed up.

Are Robotics and Artificial Intelligence the Same Thing? The first thing to clarify is that robotics and artificial intelligence are not the same thing at all. In fact, the two fields are almost entirely separate. A Venn diagram of the two would look like this:



I guess that people sometimes confuse the two because of the overlap between them: Artificially Intelligent Robots.

To understand how these three terms relate to each other, let's look at each of them individually.

What Is Robotics?

Robotics is a branch of technology which deals with robots. Robots are programmable machines which are usually able to carry out a series of actions autonomously, or semi-autonomously.

In my opinion, there are three important factors which constitute a robot:

1. Robots interact with the physical world via sensors and actuators.
2. Robots are programmable.
3. Robots are usually autonomous or semi-autonomous.

Robots are "usually" autonomous because some robots aren't. Telerobots, for example, are entirely controlled by a human operator but telerobotics is still classed as a branch of robotics. This is one example where the definition of robotics is not very clear.

It is surprisingly difficult to get experts to agree exactly what constitutes a "robot." Some people say that a robot must be able to "think" and make decisions. However, there is no standard definition of "robot thinking." Requiring a robot to "think" suggests that it has some level of artificial intelligence.

However you choose to define a robot, robotics involves designing, building and programming physical robots. Only a small part of it involves artificial intelligence.

What Is Artificial Intelligence?

Artificial intelligence (AI) is a branch of computer science. It involves developing computer programs to complete tasks which would otherwise require human intelligence. AI algorithms can tackle learning, perception, problem-solving, language-understanding and/or logical reasoning.

AI is used in many ways within the modern world. For example, AI algorithms are used in Google searches, Amazon's recommendation engine and SatNav route finders. Most AI programs are not used to control robots.

Even when AI is used to control robots, the AI algorithms are only part of the larger robotic system, which also includes sensors, actuators and non-AI programming.

Often — but not always — AI involves some level of machine learning, where an algorithm is "trained" to respond to a particular input in a certain way by using known inputs and outputs. We discuss machine learning in our article Robot Vision vs Computer Vision: What's the Difference?

The key aspect that differentiates AI from more conventional programming is the word "intelligence." Non-AI programs simply carry out a defined sequence of instructions. AI programs mimic some level of human intelligence.

What Are Artificially Intelligent Robots?

Artificially intelligent robots are the bridge between robotics and AI. These are robots which are controlled by AI programs.

Many robots are not artificially intelligent. Up until quite recently, all industrial robots could only be programmed to carry out a repetitive series of movements. As we have discussed, repetitive movements do not require artificial intelligence.

Non-intelligent robots are quite limited in their functionality. AI algorithms are often necessary to allow the robot to perform more complex tasks.

Let's look at some examples.

Example: Non-Artificially Intelligent Cobot

A simple collaborative robot (cobot) is a perfect example of a non-intelligent robot.

For example, you can easily program a cobot to pick up an object and place it elsewhere. The cobot will then continue to pick and place objects in exactly the same way until you turn it off. This is an autonomous function because the robot does not require any human input after it has been programmed. However, the task does not require any intelligence.

Example: Artificially Intelligent Cobot

You could extend the capabilities of the cobot by using AI.

Imagine you wanted to add a camera to your cobot. Robot vision comes under the category of "perception" and usually requires AI algorithms.

For example, say you wanted the cobot to detect the object it was picking up and place it in a different location depending on the type of object. This would involve training a specialized vision program to recognize the different types of object. One way to do this is using an AI algorithm called Template Matching, which we discuss in our article How Template Matching Works in Robot Vision.

Conclusion

As you can see, robotics and artificial intelligence are really two separate things. Robotics involves building robots whereas AI involves programming intelligence. Software robot" is the term given to a type of computer program which autonomously operates to complete a virtual task. They are not physical robots, as they only exist within a computer. The classic example is a search engine webcrawler which roams the internet, scanning websites and categorizing them for search. Some advanced software robots may even include AI algorithms. However, software robots are not part of robotics.

References

- [1] J.J. Craig, *Introduction to Robotics*. Reading, MA: Addison-Wesley, 2nd ed., 1989.
- [2] L. Zhenyu, C. Yinglin, and Q. Daokui, "Research on robot calibration," *Robot*, vol. 24, no. 5, pp. 447–450, 2002.
- [3] S. Lei, L. Jingtai, S. Weiwei, W. Shuihua, and H. Xingbo, "Geometry-based robot calibration method," in *Proc. IEEE Int. Conf. Robotics and Automation*, pp. 1907–1912, 2004.
- [4] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," *Int. J. Robot. Res.*, vol. 5, no. 1, pp. 90–98, 1986.
- [5] J.F. Canny, *The Complexity of Robot Motion Planning*, Cambridge, MA: MIT Press, 1988.
- [6] J.T. Schwartz and M. Sharir, "On the 'piano movers' problem: 1. The case of two-dimensional rigid polygonal body moving amidst polygonal barriers," *Commun. Pure Appl. Math.*, vol. 36, pp. 345–398, 1983.
- [7] J. Bobrow, S. Dubowsky, and J. Gibson, "Time-optimal control of robotic manipulators along specified paths," *Int. J. Robot. Res.*, vol. 4, no. 3, pp. 3–17, 1985.
- [8] J.J.E. Slotine and W. Li, *Applied Nonlinear Control*. Upper Saddle River, NJ: Prentice-Hall, 1991.
- [9] K.J. Aström and K.B. Wittenmark, *Adaptive Control*. Reading, MA: Addison-Wesley, 1989.
- [10] B. Siciliano and L. Villani, *Robot Force Control*. Norwell, MA: Kluwer, 1999.
- [11] P.C. Watson, "Remote center compliance system," U.S. Patent No. 4098001, Jul. 1978.
- [12] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2003.
- [13] R.S. Michalski, J.C. Carbonell, and T.M. Mitchell, *Machine Learning*, Palo Alto, CA: Tioga, 1983.
- [14] S.G. Tzafestas, "Fuzzy systems and fuzzy expert control: An overview," *Knowledge Eng. Rev.*, vol. 9, no. 3, pp. 229–268, 1994.
- [15] G. Niemeyer and J.J.E. Slotine, "Telemanipulation with time delays," *Int. J. Robot. Res.*, vol. 23, no. 9, pp. 873–890, 2004.
- [16] B. Davies, *A Discussion of Safety Issues for Medical Robots*, In R. Taylor, S. Lavalée, G. Burdea, and R. Moesges, Eds. *Computer-Integrated Surgery*. Cambridge, MA: MIT Press, pp. 287–296, 1996.

- [17] M. Blackwell, C. Nikou, A. DiGioia, and T. Kanade, "An image over-lay system for medical data visualization," *Med. Image Anal.*, vol. 4, pp. 67–72, 2000.
- [18] R. Kumar, P. Berkelman, Gupta, A. Barnes, P.S. Jensen, L.L. Whitcomb, and R.H. Taylor, "Preliminary experiments in cooperative human/robot force control for robot assisted microsurgical manipulation," in *Proc. IEEE Int. Conf. Robotics Automation*, pp. 610–617, 2000, San Francisco, CA.
- [19] G.H. Ballantyne and F. Moll, "The da vinci telerobotic surgical system: The virtual operative field and telepresence surgery," *Surg. Clin. North Amer.*, vol. 86, no. 6, pp. 1293–1304, 2003.
- [20] J. Marescaux and F. Rubino, "The zeus robotic system: Experimental and clinical applications," *Surg. Clin. North Amer.*, vol. 86, no. 6, pp. 1305–1315, 2003.

Faculty assignment and timetabling using optimisation

MA Nang Laik

Senior Lecturer, School of Business
Singapore University of Social Sciences
Singapore
E-mail: nlma@suss.edu.sg*

SEN Prabir

Director Research, STATGRAF Research
Vancouver, Canada
Email: sen.prabir@gmail.com

Abstract— Resource planning in university is a very hard management science problem. Faculty members are expensive resource that a university needs to utilize them efficiently and deploy them effectively for courses that they can teach. In this paper, we focus on one of the most important problems in the universities – the academic calendar which comprised of faculty-course assignment, course scheduling and timetabling. We propose an innovative two-steps approach to solve the problem using mathematical models to optimize the resource allocation while satisfying the faculty preferences. We also showcase using a real-world example how this problem is solved easily and solution improves the productivity of the staff and enhances the satisfaction of faculty.

Keywords—*faculty assignment, timetable, university, minimize cost, class schedule*

I. INTRODUCTION

Universities today are very complex and they serve thousands of students annually ranging from undergraduates, post-graduates to executives training programmes. The mission of the university in this study is to provide lifelong education, equipping learners to serve society in a positive manner. For universities, faculty members (or “faculty”) are the most expensive resources. Faculty involves in course development, assessment writing, and course delivery and grading. In addition, they also need to support other administrative tasks such as student recruitment, information sharing session, applied project supervision and industrial relationship. Some of them focus on research studies that require collaboration with the industry as an integral part of the faculty duties to keep in touch with the industry, understand the market needs and enhance the quality of classroom teaching. In the meantime, if there is an internship opportunity from the industry, students attend interviews and perform short-term internship in the company, for say ten weeks, or, if long-term, for six months.

There are two semesters in the university, one run from August to December which is referred as “term 1” and January to June as “term 2”. Before the start of each semester, the administrative team needs to forecast the demand for each course and determine the required number of sessions for each course. Course is a series of classes, on a particular subject. A faculty may teach one session of the course, or teach multiple session of the same course. University offers some courses only

once a year and others in every term. Some of the courses are foundation courses and, thereby the demand for such courses is relatively higher than the elective course, which may only have a single session. We also have a minimum class size of 20 and maximum class size of 50. If the demand falls below 20, university no longer offers the course to avoid sub-optimal use of resources. We also focus on the class interactivity thus we limit the number of students in the class to be greater than or equal to 20 but less than or equal to 50.

Faculty includes full-time employees of the university who hold Ph.D or equivalent doctorate degree. Faculty assignment problems is a non-trivial problem as the faculty are not homogenous. They can only teach those courses that are in their area of experts or fields of study, and not all other courses, as the mismatch skills and knowledge which is not ideal for both faculty and students. Some professionals or instructors from other institutions or industry who are associated with the university (e.g., by teaching some courses or supervising students) but do not hold professorships may be appointed as adjunct faculty or associates. There is a large pool of associates who teach courses in their area of expertise. University deploys them for unfulfilled courses by the full-time faculty. In any case, the associates are part-time staffs on contract and they are not entitled to receive any medical and HR benefits like the full-time faculty.

The faculty-course allocation (i.e. faculty assignment) is an important decision making process for the associate dean of the school with his administrative team need to make sure that all the course demands are met. All the courses are assigned to the faculty members (Full-time and part-time) who can teach the course well. The team also needs to ensure that all the full-time faculty members are fully-utilized before engaging the associates. This is done manually purely based on the historical teaching record and most of the times, faculty preferences are not taken into consideration. This raises a lot of concerns and dissatisfaction among the faculty members. The administration team also need to schedule the course to various timeslots based on the availability of the classroom. This is timetabling problem.

The stated problem, faced by associate dean and administrative team every semester, is done manually now. It is mundane and involve a lot of man-hours to complete the task. Thus, we have proposed a decision support system which can plan the faculty

assignment and timetabling automatically with minimal human intervention taking into consideration of operational constraints.

In the next few sections, we discuss the literature review; the important decision making process at one of the private universities in Singapore; data analysis to identify the issue of using the “gut-feel” and experience to do the resource planning at the university. We propose two-steps optimization model for faculty assignment and timetabling to meet the required demand and run the programme. The approach yields better results and improves the faculty satisfaction level as they are happy with the course allocation and schedule. We also present some computational result and performance of the models. Finally, we discuss about limitation of the model proposed and future work to be done as a conclusion.

II. LITERATURE REVIEW

Generating academic calendar for a university is a very challenging and time consuming task due to diverse demands, faculty preference and availability, limited time slots and class rooms’ requirement. Although, there are other assignment problems in a university, academic calendar problem is the most frequent and scheduling classes is the most challenging one. The academic calendar problem consists two kinds of assignment problems. One is the faculty assignment: to determine which faculty is assigned to the course based on their expertise and preference. And, the other is course scheduling: to determine the correspondence between class-timeslot and classrooms. The context and complexity of the assignment problems are dependent on the relevant university systems and various approaches have been proposed for the problem and solved by various researchers [1, 2, 3, 4, 5, 6 and 7].

Adewumi, Sawyerr, and Ali [2] addressed lecturer scheduling at a Nigerian University, and uses an iterative process to generate schedules based on the degree of violation of hard constraints. Daskalaki and Birbas, [5] developed a two-stage procedure for a department providing structured curricula for well-defined groups of students. The procedure includes a relaxation approach for computationally heavy constraints, and sub-problems to obtain timetables for each day of the week. Derigs and Jenal [7] described a Genetic Algorithm – based system for professional course scheduling using strategies such as pre-assigning subsets of courses. Dinkel Mote and Venkataramanan [9] used a network-based model considering the dimensions of faculty, subject, time, and room for the College of Business Administration at Texas A&M University. Other articles describing heuristic approaches to course scheduling in university environments include [4, 6, 8 and 10].

Fong, Asmuni, McCollum, McMullan and Omatu [11] proposed a new hybrid method which is a combination of a great deluge and artificial bee colony algorithm (INMGD-ABC) to solve the university timetabling problem. Artificial bee colony algorithm (ABC) is a population based method that has been introduced in recent years and has proven successful in solving various optimization problems effectively. However, as

with many search based approaches, there exist weaknesses in the exploration and exploitation abilities which tend to induce slow convergence of the overall search process. Therefore, hybridization is proposed to compensate for the identified weaknesses of the ABC.

Gunawan, Ng and Poh [12] developed a mathematical model to solve teacher assignment and course scheduling for a master course. An initial solution is obtained by a mathematical programming approach is based on Lagrangian relaxation. This solution is further improved by a simulated annealing algorithm. The proposed method has been tested on instances from a university in Indonesia, as well as on several randomly generated datasets, and the corresponding computational results are reported.

Hinkin and Thompson [13] considered integrated teacher assignment and course scheduling at a university in Indonesia, and used a heuristic based on Lagrangian relaxation. The models were solved in phases using CPLEX[8]. The authors developed a computer program to automate the scheduling process, considering conflicts among core required courses, and among electives within areas. The program was used by an administrator in the student services office.

Koide [14] developed a prototype system for the examination proctor assignment in Konan University by reference to the mathematical modeling by [16]. They focused on the proctor assignment and the target model considered some different types of constraints with respect to workload in a day from the constraints in [16] model. A mixed integer programming model was proposed and an optimal solution was derived through CPLEX [8], commercial optimization software. The resulting assignment sounded acceptable for the registrar staffs nevertheless some additional practical conditions were neglected for simplification of the mathematical model. This study extends the previous model and discusses the usefulness of the system for system users in the practical assignment task. The timetabling problem is generally large, highly constrained, and solution by exact optimization methods is difficult [15].

Onouchi, Uchigaito Sasaki M. [16] studied timetabling problem for final examinations in their university. The problem was solved in two stages; examination timetabling and classroom assignment were conducted in the first stage and proctor assignment in the second stage. The problems in both stages were formulated as mixed integer programming (MIP) and solved by using commercial optimization software. The authors proposed meta-heuristic approach as it is more easily comprehensible for system users.

Tim Roughgarden [17] discussed the problem of optimizing the performance of a system for the concrete setting of scheduling “centrally controlled” jobs and formulated this goal as an optimization problem via Stackelberg games, games in which one player acts a leader (here, the centralized authority or academic dean interested in optimizing system performance)

and the rest as followers (the faculty members). The problem is then to compute a strategy for the leader (a Stackelberg strategy) that induces the followers to react in a way that (at least approximately) minimizes the total latency in the system.

Based on the literature review, we have identified that faculty assignment and course-scheduling or timetabling is a very complex yet critically important to the university. Most researchers develop mathematical models and solve it using search algorithms or heuristics to achieve the goal. The contribution of our paper is as followed, we develop a two-steps approach to solve the faculty assignment and timetabling sequentially and finally use the output from both models to get the faculty schedule. This approach reduces the computation time tremendously as the number of variables are greatly reduce using our approach.

III. PROBLEM DESCRIPTION

The university discussed here is a new university which has just started operating recently. In each year, for performance appraisal, university ask faculty members to put in their preferences (first, second, third choices) in term of courses to teach. Full-time faculty members teach six classes in an academic year as per the employment contract; however, they may get some reduction in teaching load if they involve in course development or significant administrative work to support the programme. Their performances are measured by substantive contributions to the learning of their students and to their field, as well as to make service contributions to their field and the university. Therefore, student's feedback along with effectiveness of course delivery, quality of course development and effectiveness of mentoring students are essential for their career development and mentoring to ensure accountability and equity across the faculty. In essence, their performance indicators are of two kinds: 1) those that denote scholarly activities relevant to a performance area and 2) those that provide service support of the quantity and quality of content delivery activity in a performance area. Neither the number of activities nor the number of supporting services necessarily indicates a high (or low) quality of performance; instead, consider a combination of quantitative and qualitative elements when evaluating performance.

The adjunct faculty members are part-time staff and they need to fill in a form to indicate the courses that they are eligible to teach. Due to some human resource policy and regulation, adjunct faculty members teach no more than two classes in a year. By the end of term 2, the administrative team has the estimated number of students enrolled for each course before the new academic year begins in August. Based on the student enrollment, they compute the number of sessions required for each course. Each class needs a minimum class size of 20 students and a cap of 50 due to the classroom capacity. If there are 120 students signed up for the course, for example, then the total number of sessions required is 3 and they try to balance the number of students in each class. Faculty are assigned to courses

based on their expertise and the current objective is to ensure that all the courses demand are being fulfilled taking into considering of faculty preference. It is a multi-faceted decision making process for which a system will be required to solve it effectively to assign faculty to courses that they are able to teach and utilize them effectively.

After the faculty assignment, the administrative team works on scheduling the classes. There are only fifteen time-slots available in a week. And, there are three time-slots daily, we referred as "t1 or morning" for 08:30am to 11:30am, "t2 or afternoon" for 12:00pm to 3:00pm and lastly "t3 or evening" for 3:30pm to 6:30pm. Since there are a limited number of classrooms available in the university, classes are assigned to a timeslot, based on classroom availability. Therefore, one of the objectives is also to schedule all the courses to a timeslot and to an available classroom. At the end of this process, the timetable for all the course which denotes day of week, timeslot and the classroom, will be available.

Once courses allocated and classes scheduled, the faculty members who teach the course are assigned to the schedule. If a member of a faculty has conflict with the allocated schedule, the member is allowed to internally swap with another faculty member within the course. A faculty member cannot teach more than one course at the same time-slot of the day. The final output is the faculty teaching schedule.

The university would like to harness on the power of analytics, to develop a decision support system which automatically allocate, assign and schedule in that sequential order. The proposed solutions aide them in their decision making and achieve an optimal outcome.

IV. MODELING & COMPUTATIONAL RESULT

We model the faculty assignment problem as an integer programming (IP) model. The objective for the model is to minimize the cost. The cost of allocating a course to a faculty is partially based on the preferences list. Each faculty member is asked to submit three preference courses each year. The model aims to assign the course to the faculty's preference as far as possible so as to increase the satisfaction level of the staff. We assign lower cost if the course is in the preference list. The lowest cost is assigned to the first preference, which is followed by second and third choices. Example: we may assign 100 to first choice, followed by 200 to second choice and 300 to third choice. Higher penalty is applied for assigning a course to a faculty member outside of the preference list. It is also more expensive to engage an adjunct than the full-time faculty, therefore treat them separately from now on.

Let i be the number of courses, $i = 1, 2, 3, \dots, n$

Let j be the number of faculty, $j = 1, 2, 3, \dots, m$

Let x_{ij} be the decision variable, where course i is allocated to faculty j

Let d_i be the demand for the course i

Let f_j be the number of sessions required to teach for faculty j
Let c_{ij} be the cost of assigning course i to faculty j

Problem P1 – Faculty assignment

$$\text{Objective: } \min \sum_{i=1}^n \sum_{j=1}^m c_{ij} x_{ij} \quad (1)$$

$$\sum_{j=1}^m x_{ij} \geq d_i, \forall i = 1, 2, 3, \dots, n \quad (2)$$

$$\sum_{i=1}^n x_{ij} \leq f_j, \forall j = 1, 2, 3, \dots, m \quad (3)$$

$$x_{ij} \in \mathbb{Z}^+ \quad (4)$$

The first equation is the objective function, we want to minimize the cost of assignment which is translated to meeting the preferences of each faculty as much as possible while meeting all the operational constraints.

The equation (2) stated that number of faculty assigned for each course must be larger than or equal to the demand for the course. Since it is a minimization problem, the system assigns the minimum classes to the faculty. Equation (3) noted that the number of classes assigned to each faculty i , must be less than the required number of teaching sessions required. For normal full-time faculty, f_j is 6 and for adjunct faculty, f_j is 2.

This mathematically model is compact. The challenge that we face during the modeling is to assign appropriate cost to c_{ij} . We have taken into consideration of faculty's preference, faculty's expertise as well as the cost of conducting a course by full-time and adjunct. The cost of conducting a course for full-time faculty also varies according to the rank. Currently there are three ranks in the university, they are namely assistant professor, associate professor and professor ranks. The teaching cost for an assistant professor is the lowest, followed by 30% increased for associated professor and 50% increased for professor. These are users input and the users can change these inputs according to university pay-structure.

After this assignment, the team initiates timetabling process by scheduling the courses to timeslots where there are available classrooms. There are some basic assumptions for this model. The class schedule/timetable for the course is the same weekly throughout the semester, which mean that if a course BA100 is scheduled on Monday 8:30am and the venue is room 4015, then the same schedule is valid for each week within a semester. The courses can be scheduled for three timeslots: (t1, t2 t3) from Monday to Friday, total of 15 timeslots, T, in a week. In this case, we try to find the best time slot to schedule classes each week according to classrooms availability. The schedule for a class remains the same for the whole semester, unless they fall on a public holiday and, in that case, we find a make-up class on Saturday of the same week. Since the university is expanding, the classroom facilities become scarce resource. We schedule a course at a particular timeslot and to an available classroom. Thus, the objective is to schedule all the courses to timeslots and available classrooms. After this schedule process, then we assign faculty members who teach the course to a timetable. If a faculty member has a conflict with the scheduled timetable,

then we allow them to internally swap with another faculty member from the same course. A faculty member cannot teach more than one course at the same timeslot of the day, they should be allocated different timeslots. Timetable is a collection of timeslots for a week, for a course or for a faculty member.

Let i be the number of courses, $i = 1, 2, 3, \dots, n$

Let t be the number of timeslots available, $t = 1, 2, 3, \dots, T$

Let y_{it} be the decision variable, where course i is allocated to time t

Let d_i be the demand for the course i

Let r_t be the number of classrooms available

Let c_{it} be the cost of assigning course i to timeslot t

Problem P2 – Timetabling

$$\text{Objective: } \min \sum_{i=1}^n \sum_{t=1}^T c_{it} y_{it} \quad (5)$$

$$\sum_{t=1}^T y_{it} \geq d_i, \forall i = 1, 2, 3, \dots, n \quad (6)$$

$$\sum_{i=1}^n y_{it} \leq r_t, \forall t = 1, 2, 3, \dots, T \quad (7)$$

$$y_{it} \in \{0, 1\} \quad (8)$$

The equation (5) is the objective function, we want to minimize the cost of assigning course i to timeslot t , if we want to schedule more afternoon classes provided there are classroom available, then we can assign less cost to noon timeslot each day. It is generally true that students prefer noon classes than early morning class or evening class.

The equation (6) stated that number of timeslots allocated to each course must be larger than or equal to the number of classes required for the course. Equation (7) noted that the number of classes assigned in each time slot t , must be less than the classroom available. Finally, the decision variable, y_{it} is a binary variable; it is assigned 1 if course i is assigned to time t or 0 otherwise. Assuming, for C1 courses there are 3 classes weekly. The weekly class schedule can be on t1 (Monday morning), t11 (Thur noon), t14 (Fri noon).

Finally, we match the faculty assignment to the course timetable using some basic rules and we get the faculty schedule. We ensure that one faculty can't teach two classes at the same timeslot. Otherwise, we perform some minor adjustments.

The model can be solved very quickly using Excel solver if the problem is not big or using some commercial software like IBM CPLEX Optimizer. We have used past university data and run the model for 1 year. For the 20 courses, 80 classes and 16 faculty, we are able to solve the problem in less than 30 minutes. It greatly improve the productivity of the administrative team and they just need to look at the preliminary result and made some minor adjustments when required.

We use the following example to illustrate how our approach works. Assuming we have 10 courses, 40 classes and 5 full-time faculty and 5 part-time adjunct faculty. The faculty has given their preferences as well as courses that they can teach in advance. There are 15 timeslots in a week namely, t1, t2 .. t15.

t1 represents Monday morning, t2 represents Monday noon and t15 for Friday evening. There are also only three classrooms available to book for each timeslot. What is the faculty assignment, timetable for the courses and faculty timetable? Using the above formulation and using Excel solver, we are able to solve the problem in minutes. Here are some sample output from the model.

This formulation has some limitations. The model may allocate different courses to faculty, e.g. they need to teach four different courses in a year which is neither desirable as one faculty shouldn't teach more than three different courses in a year, nor optimal for the program. But the occurrence of this allocation is very rare as the model try to assign the course within the faculty preference list of three choices.

Assuming that we have 10 courses, 10 faculty member to assign in 15 timeslot, the number of variables to solve all of them in one model is $10 \times 10 \times 15 = 1500$ variable. If we split the problem into two-steps approach, the total number of variables for problem P1 is $10 \times 10 = 100$ and P2 is $10 \times 15 = 150$, which is much lower than 1500. If we increases the number of courses to 100 courses, 100 faculty and 15 timeslot, the number of variables become very large and computation time increases exponentially.

Next, we are going to use a simple example to illustrate how our model work in real-world.

For a small business programme in our university, we only have 5 full-time faculty members and 5 adjuncts. We are going to offer 10 courses in the coming year and based on the past demand, we know that the number of classes we need to be offered is 40. Refer to the input [Table 1](#) below for detail. For example, C10 is the foundation course, we need to offer 10 sessions of C10 in a year, but C3 is an elective so there is only 1 session offered. Each faculty is asked to fill in a table to state their preference as well as up to 3 other differences courses that they can teach based on their expertise. Adjunct faculty are asked to also fill in up to 3 courses that they can teach. These form the input ([Table 1](#)) for our mathematically model.

Table 1: Input for example:

Course demand:

course	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
demand	3	2	1	2	6	3	4	5	6	8

Faculty preferences and courses they can teach:

Resources	Courses you can teach		
	1 st	2 nd	3 rd
R1	C5	C3	C2
R2	C1	C4	C5
R3	C3	C8	C6
R4	C6	C8	C7
R5	C2	C1	C5

Resources	Courses that adjuncts can teach		
	C1	C3	C5
R6	C1	C3	C5
R7	C2	C4	C8
R8	C1	C5	C8
R9	C3	C6	C9
R10	C5	C7	

Number of classrooms available: 3

Timeslots: t1, t2, t3, ... , t15

Table 2: Output from P1

Resources	Courses assigned
R1	C2(1), C5(1), C7(4)
R2	C1(2), C9(4)
R3	C4(2), C6(1), C8(3)
R4	C6(2), C8(2), C10(2)
R5	C2(1), C5(5)
R6	C3(1), C10(1)
R7	C10(2)
R8	C1(1), C10(1)
R9	C9(2)
R10	C10(2)

We can derive the courses that each faculty needs to teach from the table above and share with the faculty who are teaching the courses – refer to [Table](#). We can verify that we assign all the courses and sessions to faculty refer to their preference list. The faculty are also aware that the system will try its best to find the best match but it may not be able to fulfill all their wishes as the most important problem is to satisfy the operational constraints to ensure that all courses are assigned at the end of this stage with minimum cost.

After we have run **P1** model, we continue with the class timetabling problem **P2**. For timetabling, we need to assign a class room and timeslot for each class subjected to room availability. The output of this model P2 is a weekly timetable for all the courses and their respective classroom. A faculty teaches a course, which consists of multiple session of weekly classes over the whole semester. A class timetable denotes day of week and timeslot in the designated class room weekly, thus we will use class and course interchangeable for this purpose.

In this example, there are only 3 class room available at each time slot and there are altogether 15 time slots available each week. The output from P2 is shown in

Table 2.

Using the output from **P1** and **P2**, we need to schedule the faculty timetable. This is done using a simple rule. One faculty can only teach one class at any point in time, if there is a clash, we swap the time table with another class.

R10 t7,t8

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
t1	0	0	0	0	0	0	0	0	1	1
t2	0	0	0	0	1	0	0	0	1	1
t3	0	0	0	0	0	0	0	1	1	1
t4	0	0	0	0	0	0	0	0	1	1
t5	0	0	0	0	1	0	1	0	0	1
t6	0	0	0	0	0	0	0	1	1	1
t7	0	0	0	0	0	0	0	0	1	1
t8	0	0	0	0	0	1	0	1	0	1
t9	0	1	0	0	1	0	0	1	0	0
t10	0	0	0	0	0	0	1	1	0	0
t11	0	0	0	1	1	1	0	0	0	0
t12	0	0	0	0	1	1	1	0	0	0
t13	1	0	0	1	0	0	1	0	0	0
t14	1	0	0	0	1	0	0	0	0	0
t15	1	1	1	0	0	0	0	0	0	0

Table 2: Output from P2 - (Timetabling)

Table 3: Matching the faculty assignment with schedule:

R1	C2 (1)	C5 (1)	C7 (4)
	t9	t2	t5,t10,t12,t13
R2	C1 (2)	C9 (4)	
	t13, t14	t1, t2, t3, t4	
R3	C4 (2)	C6 (1)	C8 (3)
	t11, t13	t8	t3, t6, t9 (t8 clash so change to t9)
R4	C6 (2)	C8 (2)	C10 (2)
	t11, t12	t8, t10	t1, t2
R5	C2 (1)	C5 (5)	
	t15 (t9 can't be assigned)	t5, t9, t11, t12, t14	
R6	C3 (1)	C10 (1)	
	t15	t3	
R7	C10 (2)		
	t4, t5		
R8	C1 (1)	C10 (1)	
	t15	t6	
R9	C9 (2)		
	t6,t7		
	C10 (2)		

In this model, the associate dean of the university and the system (responsible for assigning the centrally controlled resources and interested in optimizing welfare of faculty and students) act as a leader, in that it may hold its faculty assignment (its strategy) fixed while all other agents (the followers) react independently to the leader's strategy, reaching a Nash equilibrium relative to the leader's strategy. As in game theory, this Stackelberg games, and resulting Stackelberg equilibria of the model for the faculty assignment is induced by a strategy that is precisely the optimal assignment of all of the courses.

V. CONCLUSION

In conclusion, we propose an innovative method to solve the faculty assignment and timetabling problems for the university using two-steps approach. The problem is solved in a short running time of 30 minutes and it assists management team in term of avoiding tedious and manual planning. It also takes faculty preference into consideration for the course, thus the outcome sharply reduces the conflict and, thereby, improves productivity and yields higher satisfaction of the faculty members. By using this approach, we also reduce the number of variables and errors in runtime, thus it can be used for small and medium size university for their resource planning project.

The limitation of our model is as two folds. First, we assume that the demand for the course is known and it does not vary too much since the courses starts. Otherwise, it may not be economically viable to run the course if the number of students in the class is less than the breakeven number. We can resolve this during the operations by cancelling the course, if we realize the actual number of sign up is lower than the breakeven number. However, this can only be done before the start of the course. Once the course starts, even if some students drops off half-way and the number of students in the class falls below the breakeven number, the course still continues and remains sub-optimal. This is to avoid any disruption in students' study plan or delay their graduation. Second, we assume that we always have enough adjuncts or faculty members to teach a course. If there is a mismatch, the model might not find a feasible solution. This can be overcome by ensuring enough faculty to teach the course. Otherwise, the course may not be offered until we find a faculty to teach it.

This approach can also be used in other related industry where resource assignment or planning is required. In finance, we can use our model to solve the allocation of financial advisor to potential investor and scheduling appointment. In healthcare, we can deploy our model to schedule doctor and patient face-to-face appointment or operation theater scheduling based on doctor's preference time slot and operation theater's availability.

For the next step forward, we want to explore solving the problem as one-step approach. We have acknowledged that this

problem might be too big or complex to be solved using the mathematical model to find the optimal solution within a reasonable timeframe. Thus, we will develop some heuristics to solve the real-world problem in the near future.

REFERENCES

- [1] Abdullah S, Turabieh H. "On the use of multi neighbourhood structures within a tabu-based memetic approach to university timetabling problems", *Information Sciences* 2012; **191**:146-68.
- [2] Adewumi, A. O., Sawyerr, B. A., and Ali, M. M. "A heuristic solution to the university timetabling problem", *International Journal for Computer-Aided Engineering and Software* 2009, Vol. 26, No. 8, 972-984.
- [3] Beligiannis GN, Moschopoulos CN, Kaperonis GP, Likothanassis SD. "Applying evolutionary computation to the school timetabling problem: The Greek case", *Computers & Operations Research* 2008; Vol **35**:1265-80.
- [4] Boronico, J. "Quantitative modeling and technology driven departmental course scheduling", *Omega*, Vol. 28, Issue 3, pg. 327.
- [5] Daskalaki, S., and Birbas, T. "Efficient solutions for a university timetabling problem through integer programming", *European Journal of Operational Research* 2015, Vol. 160, Issue 1, pg. 106-120.
- [6] Daskalaki, S., Birbas, T., and Housos, E. "An integer programming formulation for a case study in university timetabling", *European Journal of Operational Research* 2004, Vol. 153, Issue 1, pg. 117.
- [7] Derigs, U., and Jenal, O. "A GA-based decision support system for professional course scheduling at Ford Service Organisation", *OR Spectrum* 2005, Vol. 27, Issue 1, pg. 147.
- [8] Dimapoulou, M., and Miliotis, P. "An automated university course timetabling system developed in a distributed environment: A case study", *European Journal of Operational Research* 2004, Vol. 153, Issue 1, pg.136.
- [9] Dinkel, J. J., Mote, J., and Venkataramanan, M. "An efficient decision support system for academic course scheduling", *Operations Research* 1989, Vol. 37, Issue 6, 853-864.
- [10] Ferland, J., and Fleurent, C. "SAPHIR: A decision support system for course scheduling", *Interfaces* 1994, Vol. 24, Issue 2, pg. 105.
- [11] Fong CW, Asmuni H, McCollum B, McMullan P, Omatu S. "A new hybrid imperialist swarm-based optimization algorithm for university timetabling problems." *Information Sciences* 2014; **283**:1-21.
- [12] Gunawan, A., Ng, K. M., and Poh, K. L. "A hybrid Lagrangian relaxation and simulated annealing method for the course timetabling problem", *Computers and Operations Research* 2012, Vol. 39, 3074-3088.
- [13] Hinkin, T. R., and Thompson, G., M. "SchedulExpert: Scheduling courses in the Cornell University School of Hotel Administration", *Interfaces* 2002, Vol. 32, Issue 6, 45-57.
- [14] Koide T. "Prototype system development for examination proctor assignment problem using mixed integer programming." *Proceedings of the World Congress on Engineering and Computer Science* 2014; Vol. 2:961-4.
- [15] Mirrazavi, S. K., Mardle, S. J., and Tamiz, M. "A two-phase multiple objective approach to university timetabling utilising optimisation and evolutionary solution methodologies", *The Journal of the Operational Research Society* 2003, Vol. 54, Issue 11, 1155-1166.
- [16] Onouchi O, Uchigaito T, Sasaki M. "Examination timetabling problem in universities.", *Abstracts book of ORSJ spring meeting 2013*; 66-7 (in Japanese).
- [17] Tim Roughgarden, *Stackelberg scheduling strategies*, *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, p.104-113, July 2001, Hersonissos, Greece

A Novel and Robust Retroareolar Region Detection in Mammography Images

Dr. P Pedda Sadhu Naik¹

¹Professor & HOD

¹Dr. Samuel George Institute of Engineering &
Technology, Markapur, AP,

Dr T Venugopal²

²Professor & HOD

²JNTUH College of Engineering, Jagityala,
Telangana.

Abstract – Now a day, most of the women has been suffering from Breast cancer. For assessing and segmenting the breast cancer, Mammographic image analysis is a vital tool. Most of the researchers in the literature have shown that the Retroareolar (RA) region of the breast is an important part for detecting the cancer there by the performance of algorithms can be enhanced. However, all the conventional RA region detection algorithms failed to show the reliability and most of them are manual segmentation algorithms. Here, in this proposal we designed and implemented a novel and automated RA region detection in mammography images. Our proposed frame work has been divided into three steps to detect the RA region effectively. Simulation analysis has shown that our proposed algorithms outperform the traditional approaches.

Keywords: *Mammography, RA region, breast segmentation, Hough transformation and morphological operations*

I. INTRODUCTION

The most common female cancer in the world is Breast cancer with an averaged 1.67 million new cases of cancer have been diagnosed in 2015. While the age adjusted incidence rates of breast cancer in India is lesser than the foreign countries, because of the huge population the load of breast cancer is high, about 1/3rd in urban and 1/9th in rural regions. The lack of population screening in India undoubtedly contributes to this statistic but more importantly, so do lifestyle, reproductive and dietary factors. There need to be systematic efforts at researching, preserving, and promoting those factors that “protect” Indian women from breast cancer. Typically, there are four main types of breast cancer: ductal carcinoma in situ (DCIS) where the cancer is confined within the ducts of the breast, lobular carcinoma in situ (LCIS) where the cancer is confined within the lobules or glands of the breast, invasive ductal carcinoma (IDC), and invasive lobular carcinoma (ILC). IDC and ILC refer to the type of breast cancer where the tumor has spread from the ducts or lobules it originated from, respectively, into the surrounding tissue of the breast. Other less common breast cancers include medullary carcinoma, mucinous carcinoma, Paget’s disease of the nipple,

Phyllodes Tumor, and tubular carcinoma. Breast cancer is grouped into stages which indicate the invasiveness of the disease. There are four stages—I, II, III, IV—defined by the American Joint Committee on Cancer based on a combination of tumor size, lymph node involvement, and presence or absence of distant metastasis. There is also a more general classification: early/local stage where tumor is confined to the breast, late/regional stage where cancer has spread to the surrounding tissue or nearby lymph nodes, and advanced/distant stage where cancer has spread to other organs beside the breast. There has been a decline in breast cancer mortality rates of about 2.3% over the last decade due to improved screening techniques leading to earlier detection, increased awareness, and improved treatments.

II. RELATED WORK

Diagnostic of breast cancer in primary stages is vital for enhancing the full recovery probability and for mitigating the rate of associated mortality. In present days, breast cancer’s early detection has been done by mammography screening, which is the most widely utilized, effective and low cost technique [1]. For detecting the breast cancer and their assessment can be done by computerized mammographic image analysis. In order to study parenchymal patterns [2], image-based biomarkers have been applied by various researchers on particular regions of interest (ROI). In this orbit, conventional works have exhibited that the image based biomarkers association is superior in the zone immediately behind to the nipple, namely the Retroareolar (RA) region [3].

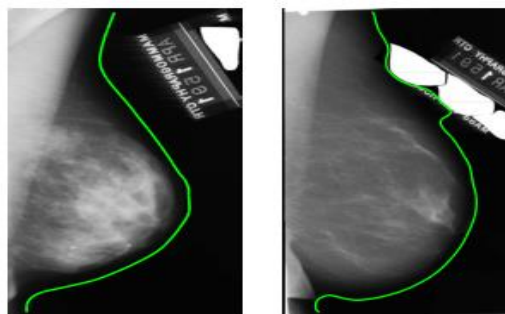


Fig.1 Example of wrong segmentation

Later on, some amount of research has been done for the parenchymal tissue analysis in the RA region [4]. In spite of its promising outcomes, the major constraint of the said works is the requirement of human interaction for manual RA region segmentation. First most, subjectively human-composed ROIs imply certain restrictions in terms of reproducibility and scalability (e.g. the application of these methods to large datasets). Secondly, most of literature works used fixed and squared ROIs, which might be a problem for adapting shapes and sizes of wide breast verities. To address the above mentioned problems, here we proposed a novel and robust methodology for automatic RA region detection in mammographic images. We considered the input breast image geometry for automatic adjustments to diverse shapes and sizes. For this, we have built upon recent implementations in the anatomical coordinate systems creation in mammographic images [5], [6] and [7].

III. PROPOSED METHODOLOGY

Segmentation of breast is arguably the very primary pre-processing step in the algorithms of mammographic image analysis. Here, we performed the segmentation in two steps: scanning artifacts removal and contour detection. Tape artifacts are markings left by tapes, or other shadows that appear as horizontal or vertical running strips in an image. Since these are straight lines, the algorithm 2 has used to segment the foreground and for detecting the artifact lines.

Algorithm 1: Breast Segmentation

Case 1: RMLO view

Case 2: LCC view

Step 1: Select and read an input mammographic image

Step 2: Find out whether the side is left (L) or MLO using string comparison

Step 3: If the side is left then flips it horizontally

Step 4: Now, detect the breast foreground using algorithm 2

Step 5: Detect the chest walls in the case of MLO using algorithm 3

Step 6: Finally, overlay segmented mask on input breast image

For the second step, contour detection in breast, a statistical chest wall segmentation technique was used in this proposal work. By using algorithm 3 after step 6, the post processing step will be done for obtaining the binary mask by means of morphological operators in order to remove spurious artifacts and the breast contour is smoothed.

Algorithm 2: Foreground segmentation

Input: I and binary flag

where I = Grayscale mammography image with size

of $m \times n$ and binary flag = 1 if the input is MLO view

Output: out_mask

Step 1: Initialize input parameters and find the central region

Step 2: Calculate maximum and minimum intensity values

Step 3: Find out the intensity threshold using the relative frequency and convolution

Step 4: Artifact removal with connected components labeling by keeping largest left-most cluster

Step 5: Now, find the contour points using region boundary tracing

Step 6: Finally, obtain the out_mask by using curvature analysis

Algorithm 3: Chest wall segmentation

Input: grayscale mammography image

Output: segmented chest wall

Step 1: First, crop the mammography image according to the contour

Step 2: Apply dilation and filtering for pre-processing of input mammography image and replace Step 3: the lower right corner with zeros

Step 4: Now, detect the pectoral line using Hough transformation then calculate the edge co-ordinates

Step 5: Calculate the accumulation array 'A' by quantizing the parameter spaces

Step 6: Find maximum in 'A'

IV. SIMULATION RESULTS

All the experiments have been done in MATLAB 2016b environment with 4GB RAM and Intel processor. The Digital Database for Screening Mammography (DDSM) [9] dataset has been utilized for the testing purpose, which consists nearly 2500 mammogram cases from various medical institutions in United States of America. Here in our experiments, we had considered the images with LCC and RMLO view by scaling them with factor 8 and later on converted to unit8 for enhancing the processing speed.

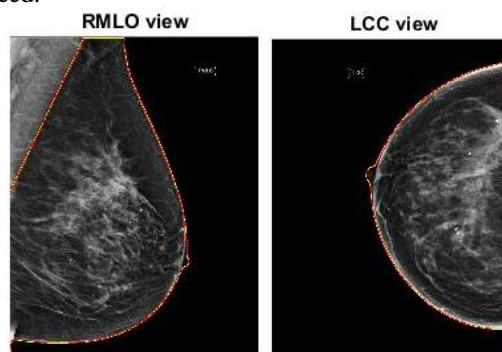


Fig.2 Segmented RMLO and LCC view mammography images

Table 1 show that the number of tested mammography images with normal and abnormal conditions. The number of correctly segmented count

also measured in the table with their accuracy in % values. Totally we tested 72 images and got an accuracy of 73.61% with our proposed methodology. Figure 2 show that the segmented output images of RMLO and LCC view images with RA region detection, which have been obtained by utilizing our proposed algorithms discussed in section III.

Table I: Performance of breast segmentation

Type	Analyzed	Correct	Accuracy
Normal	47	35	74.44 %
Cancer	25	18	72 %
Total	72	53	73.61 %

Performance of proposed and conventional RA region detection in terms of accuracy has shown in figure 3. The relation between specificity and sensitivity shown in figure 4

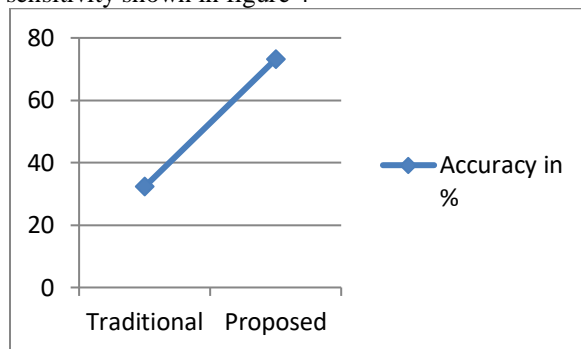


Fig.3 Performance of proposed and traditional methods

Experimental results show that the proposed method outperforms selected ROIs for the cancer detection task using proposed method.

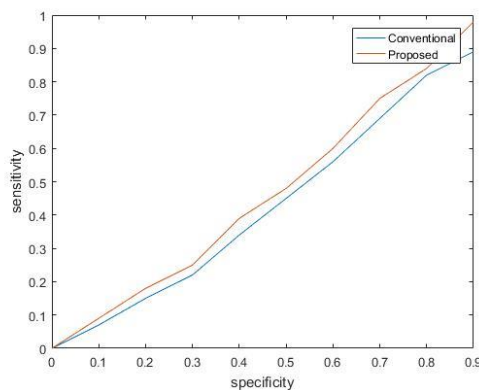


Fig. 4 Relation between specificity and sensitivity

V. CONCLUSIONS

Here, we proposed a novel and robust RA region detection in mammography images. Our proposed algorithms show the robustness of segmentation and RA region detection in terms of accuracy and sensitivity. We tested both RMLO and LCC view mammography images for showing the effectiveness of proposed algorithm. Furthermore, this can be enhanced by applying contrast enhancement

approaches for better segmentation accuracy. We can also develop RA region detection in 3D mammography images.

REFERENCES

- [1] American Cancer Society. Breast Cancer Facts and Figures 2015-2016 Atlanta:American Cancer Society, Inc. 2015.
- [2] Tahoces P. G., Correa J., Soutos M., Gomez L., Vidal J. J. Computer-assisted diagnosis: the classification of mammographic breast parenchymal patterns Physics in Medicine and Biology. 1995;40:103.
- [3] Li Hui, Giger Maryellen L., Huo Zhimin, et al. Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: Effect of ROI size and location Medical Physics. 2004;31:549– 555.
- [4] Wei Jun, Chan Heang-Ping, Wu Yi-Ta, et al. Association of computerized mammographic parenchymal pattern measure with breast cancer risk: a pilot case-control study Radiology. 2011;260:42–49.
- [5] Brandt S. S., Karemore G., Karssemeijer N., Nielsen M.. An Anatomically Oriented Breast Coordinate System for Mammogram Analysis IEEE Transactions on Medical Imaging. 2011;30:1841–1851.
- [6] Pertuz S., Julia C., Puig D.. A Novel Mammography Image Representation Framework with Application to Image Registration in 2014 22nd International Conference on Pattern Recognition (ICPR):3292–3297 2014.
- [7] Abdel-Nasser Mohamed, Moreno Antonio, Puig Domenech. Temporal mammogram image registration using optimized curvilinear coordinates Computer Methods and Programs in Biomedicine. 2016;127:1–14.
- [8] Heath Michael, Bowyer Kevin, Kopans Daniel, Moore Richard, Kegelmeyer W Philip. The digital database for screening mammography in Proceedings of the 5th international workshop on digital mammography:212–218 Citeseer 2000.

AUTHORS PROFILE

Dr. P Pedda Sadhu Naik¹ received B.Tech (CSE) Degree from JNTUniversity in 2003 and M.Tech (CS) Degree from JNTUCE Anantapur in 2007. He is received Ph D from JNTUK, Kakinada in Computer Science under the guidance of Dr T. Venu Gopal. He has 14 years of teaching experience. He joined as Assistant Professor in Dr.Samuel George Institute of Engineering & Technology, Markapur, India in 2003. Presently he is working as Associate Professor and Head of the Department of CSE. His Interested research areas are Image Processing and Soft Computing. He has life member of ISTE and IAENG. He organized various National level Technical Symposiums as Convener. He attended Various National and International Workshops and Conferences on image Processing.

Dr T Venugopal² received B.Tech (ECE) Degree from Osmania University Hyderabad in 1994 and M.Tech (CSE) Degree from JNTUH Hyderabad in 2003. He completed Ph D from JNTUH, Hyderabad in Computer Science. He is Professor, JNTUH College of Engineering, Jagityal. Jagityala District, Telangana. He has 20 years of teaching experience. He has worked in different colleges in different

levels. Presently he is working as Professor in JNTUHCE Jagityala. His Interested research areas are Image Processing and Soft Computing. He has life member of ISTE and IAENG. He published more than 50 international journals. He organized various National level Technical Symposiums as Convener. He attended Various National and International Workshops and Conferences on image Processing

Improving DBMS Security through the use of a Checksum Technique

Naim Ajlouni* Firas Ajlouni** Alaa Ali Hameed***

Istanbul Aydin University* ***
Lancashire College of Further Education**
naimajlouni@aydin.edu.tr

Abstract

We propose an approach to database security that exploits existing DBMS facilities to associate a separately maintained checksum value with critical data. Using our approach, a database's content, domain and referential integrity remain the responsibility of the DBMS, however, when critical data is manipulated checksum values are computed and stored in a separate database. Using this combination of databases, applications which access critical data can only access such data via checksum values ensuring that data is created and accessed in a secure manner.

Key words- DBMS, Security, Checksum, entity integrity, domain integrity, referential integrity and user-defined integrity.

1. Introduction

The primary responsibility for ensuring the security and integrity of a database lies with the **D**ata**B**ase **M**anagement **S**ystem (DBMS) [1]. At present, the most widely used **R**elational **D**atabases's (RDBMS's) provide several ways of ensuring data integrity at the entity, domain, referential, and user-defined levels. The latter of these levels, i.e. the user-defined level, involves constraints on the forms of update that can be performed, and such constraints are usually enforced via triggers either by the database itself or by applications that access the database.

Checksums have been the subject of research and practical application for many years. Gopalan [5] used checksums to enhance the integrity of a conventional file system via a block-level checksum computed per-block for all data blocks in a file, which is indexed by a relative block number. In such an approach, files are modified to include a set of additional references that point to the checksum blocks, and the addition of a block to a file involves the computation and storage of its checksum blocks. Crocker [6] introduced a checksum-based "Spam Detection" engine for Lotus Domino mail systems called *Block* which also exploits checksums. The system computes a checksum for each message that is classified as "Spam" and maintains the checksums in a database which is replicated between the mail system and the mail server. The database enables the server, on receipt of each message, to compute a checksum and compare it with those checksums already in the database. If a match is found, it is likely that the associated message is further "Spam". Network

Appliance Inc. technical report [7] describes a technique for reducing the volume of data transfer during Backup and Restore on UNIX platforms. The system uses a checksum to identify portions of a file that have changed since a previous and current backup. Changed blocks are identified via checksum values computed and maintained for each block. Sabarnam [8], uses checksums to detect errors in database manipulations including transactions, locks, logs, and data buffers. In this case, a checksum is added to the object being manipulated, and a checksum field is attached to the access method and also to each of the objects in associated hash buckets. A hash vector and object checksum are recomputed during the restart of a DBMS and compared with stored checksum values.

We describe an approach to database security that similarly exploits checksums. First, in Section 2, the forms of data integrity that must be maintained in a conventional RDBMS are examined. In Section 3 we present our approach to maintaining user-defined integrity using error detection and correction algorithms similar to those implemented in network protocols, i.e. the Hamming Code Protocol. [3] [4]. An implementation of the proposed approach is explained in detail in Section 4, Section 5 test and Section 6 examines properties of the implementation.

2. Data Integrity in an RDBMS

Data Integrity levels in an RDBMS can be classified into four main kinds, i.e. entity integrity, domain integrity, referential integrity and user-defined integrity [2].

Entity Integrity: Useful at the row level in a table, entity integrity ensures that a relation does not have any duplicate rows and that each

- have a unique primary key that can be defined by one or more of its attributes.
- **Domain integrity:** Values in a column in a table must be drawn from some well-defined “domain” of values. This is the simplest form of an integrity constraint which is maintained at all times and in all circumstances. In effect, a “domain” corresponds to *type* in a conventional programming language, and values in a column must be drawn from one of the available types.
- **Referential Integrity:** This is applied at the table level such that values available in one relation are available and synchronized with those in other relations. Referential integrity is enforced with a primary key and **Foreign Key** (FK) combination. A foreign key comprises one or more columns in a “child relation” whose values are synchronized with those in the PK in a “parent relation”. The FK accepts only those values that exist in the PK in order to maintain the integrity of the relation. Referential integrity is preserved when applying any **Data Manipulation Language** (DML) operations, i.e. insert, update and delete operations, via the following constraints on the application of such operations:-

1. Restricted: Disallow data modification.
2. Cascaded: Extend the data modification on parent relation to all child relations.
3. Nullified: Set the values of matching FK's in child relation's to the value NULL.

- **User-Defined Integrity:** In effect, user-defined constraints on the ways in which a database can be manipulated. Such constraints are the responsibility of the system administrator, who will administer access rights and enforce rules and regulations.

The technique described in the following section is intended to deal with user-defined integrity, a form of integrity typically enforced via triggers and constraints [9]. Our technique is intended to augment conventional approaches to user-defined integrity by exploiting conventional triggers and constraints to ensure that access to critical data is only possible via checksums.

3. Database Checksum Requirements and Prerequisites

There are several requirements that our approach must satisfy:-

- As the potential exists for “illegal” data modifications to change all the stored data, any solution should be able to both detect and correct all the modifications to data in such circumstances (completeness)

row in a relation has its own unique identifier which distinguishes it from other rows. This unique identifier is termed the **Primary Key** (PK), and each relation must

- The solution should not be restricted in the domains whose values may be used in the computation of the checksums (no special cases)
- The solution should involve no additional overhead where legal modifications are made (efficiency)
- The solution has the prerequisite that both the original database and its associated checksum database are available to all applications which access the original database, and that all such applications correctly maintain both the original database and the associated database of checksums (correctness)
- Execution time overheads and additional space requirements should both be minimal (efficiency)

3.1. Our Approach

Our approach involves the use of two databases, the first database contains conventional data of some kind and the second stores associated checksum values, i.e. a result and a remainder, that are associated with critical data in the first database. Figure (1) shows to example databases named “DB1” and “CS1”, “DB1” is a conventional database is containing values of any permissible type, e.g. integer, decimal, date including time, etc. The checksum values for critical data in “DB1” are stored in “CS1”.

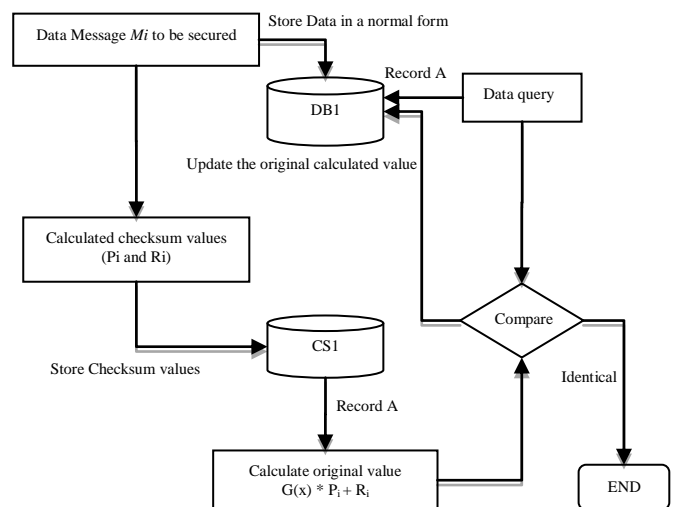


Figure (1)

Checksum values are calculated as the modulus value of a data item (the item being first converted to its ASCII value if necessary) using a divisor chosen by the system administrator.

The equation for the recalculation of the original values using the stored checksum related values:

$$D_i = G(x) * P_i + R_i. \quad 1$$

where D_i is the recalculated value, $G(X)$ is a fixed regular divisor chosen, e.g. 1028, P_i is the result of the division of M_i by $G(x)$, and R_i is the modulus of M_i and $G(x)$ are given by,

$$P_i = M_i / G(x) \quad 2$$

The modulus value is calculated using equation (3) as shown below:

$$R_i = M_i - (G(X) * (M_i \setminus G(X))) \quad 3$$

Where, M_i is assumed the i^{th} datum in a database of N data sets, \setminus is the remainder, in this case P_i and R_i are the checksum values for the i^{th} datum in a database of N sets of data, where (N is $1, 2, 3, \dots, N$), i.e.

By substituting equations (2) and (3) into equation (1) we yield

$$D_i = G(x) * M_i / G(x) + M_i - (G(X) * (M_i \setminus G(X))) \quad (4)$$

Equation (4) can be further simplified to give,

$$D_i = 2M_i - (G(X) * (M_i \setminus G(X))) \quad (5)$$

The checksum values P_i and R_i are maintained separately, e.g. in the database called "CS1" in Fig. 1., and used later to validate the integrity of the i^{th} data item stored in the original "DB1" database. When a data item in the "DB1" database is to be accessed by an application, the application must a) compute checksum values, and b) compare the computed values with those stored in the "DB1" database. If the computed and stored checksum values differ, then the associated critical data in "DB1" database is assumed to have been changed illegally, and the stored checksum values for the i^{th} data item are used to recalculate the original values and restore the database to a consistent state.

Consider, next, the worked example below:-

i. Integer Test

Let M_i denote a "critical" credit card number that is to be inserted in a "DB1" database of account information, and $G(x)$ denote a fixed regular divisor 1028. For simplicity, we assume that the initial balance in all accounts is £0.00. Table (1) below, shows a table of stored values in the Data base with their related Check sum values.

M_i	$G(X)$	P_i	R_i	D_i
7654321678945320	1028	7445838209099	548	7654321678945320

Table (1) stored data

Keeping in mind that the checksum values for P_i and R_i are stored in the "CS1" database and the value M_i is stored in "DB1" database together with other account data including the opening balance.

Assuming that subsequent to the initial insertion of the account numbered 7654321678945320 into the "DB1"

database, an illegal update takes place and the credit card number is changed to (7654321678955555). Table (2) below shows the content of the data base after the illegal modification has taken place.

M_i	$G(X)$	P_i	R_i	D_i
7654321678955555	1028	7445838209099	548	7654321678954320

Table with illegally changed M_i Value (2)

Any query issued on this account (or any other critical data) will lead to an automatically re-computing the checksum values using equation (5) which will be followed by a comparison of both recomputed and stored value of the checksum, comparing the computed values with the corresponding stored value as shown in Figure (2). Which lead to a restoration of the original value in the database to preserve database integrity?

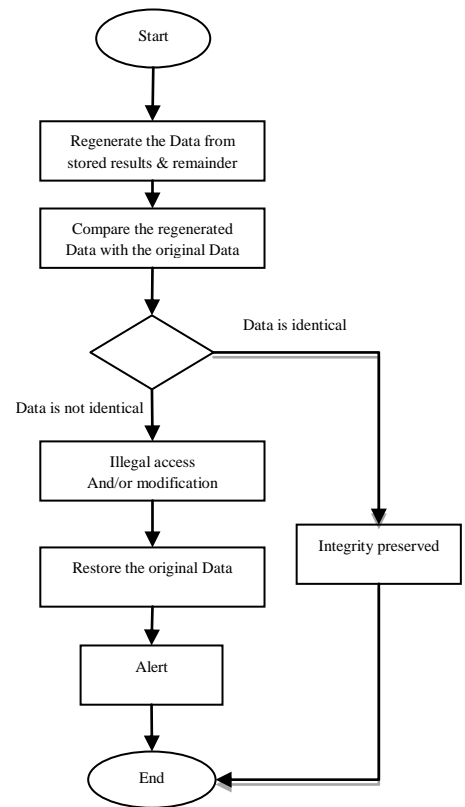


Figure (2)

If $M_i = D_i$ then the critical data is assumed to be valid, otherwise the critical data is assumed to be incorrect, but can be recovered provided only that the values of the data involved were inserted by an authorized user.

ii. Character Test

Assuming that we have a string M_i whose value is "Rob" that is to be inserted into a "DB1" database whose fixed divisor, $G(x)$, is 1028. The string is first converted to a sequence of values that correspond to its character's ASCII codes. However, all such ASCII values have the value 100

added to ensure that they are 3 digits in length. The resulting ASCII values are concatenated together to obtain a single number.

Assuming that $M_i = \text{"Rob"}$, the ASCII equivalent is 8211198, i.e. ASCII for **R** is 82, for **o** is 111 and for **b** is 98 giving 182 211 198 when the necessary 100 is added.

If $G(x)$ is 1028, the values for P_i and R_i are shown in table (3) below

Character Set	M_i	$G(X)$	P_i	R_i	D_i
Rob	182 211 198	1028	177248	254	182 211 198

Table (3) stored data

Assuming that subsequent to the initial insertion, an illegal update occurred and the name was changed to "John". Any query issued on this critical data requires the checksum values to be recomputed and compared with the stored values, i.e. the value associated with the initial insertion of "Rob", using equation (5).

Each 3 digit substring of D_i represents a character, which, when 100 is subtracted from the three digit substring yields the ASCII codes 82, 111 and 98 enabling the value in the "DB1" database to be recovered.

4. An implementation of the proposed solution

For simplicity, it is to be assumed that all access to data in a "DB1" database is protected through the use of an *application view* [12], and direct access is forbidden by the DBMS's access policy which is set by the database administrator and subject to any other third-party security system, e.g. a Firewall [10, 11].

Application triggers then perform the calculation of checksum values, i.e. result and remainder P_i and R_i respectively and their storage in the "CS1" database.

Consider, next, a customer payment system, composed of several tables, among which are the three tables shown in Figure (3) below.

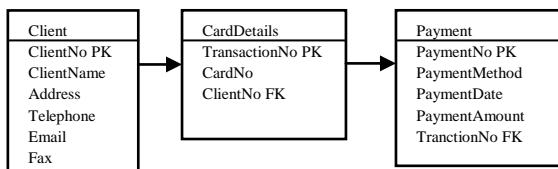


Figure (3) Case Presentation

Our interest here is to preserve the integrity and security of the following fields:

- PaymentAmount Field in the Payment table.

- CardNo Field in the CardDetails table.
- ClientName field in the Client table.

As suggested earlier, direct integrity rules provide basic integrity protection, but are unable to prevent illegal changes to values stored in a database, and similarly unable to prevent the illegal exposure of such values. In such circumstances database transactions, both "legal" and "illegal" will run normally and it will be assumed (even in the event of an illegal transaction) that the integrity of the database is preserved and uncompromised.

4.1 Validation Process

The validation process is concerned with data integrity; this process is carried out by recalculating the data related to the data checksum value using a pre saved indicators related to this checksum, and comparing those values with available data values. The validation algorithm is shown below:

```

Begin
  Identify record in Client table;
  Obtain CardNo value for defined record;
  Obtain result, remainder values for
  defined record;

  Compute Checksum
  Checksum = result X G(x) + Remainder
  Convert checksum to string;
  if checksum = 0 and
  (CardNo value = 0 or CardNo = null)
  { validate = true;}
  else if checksum ≠ 0 and CardNo =null
  { validate = false;}
  else
  {if checksum = CardNo
  {validate = true;}
  Else
  {validate = false;}
  }
End
    
```

4.2. Number implementation

This example will be applied to the CardDetails and NCalculations as shown in Figure (4):

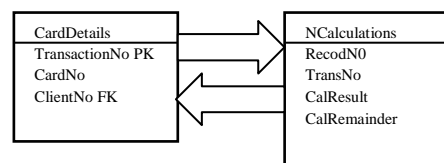


Figure (4) Presented Case

CardDetails Table: Any illegal changes to values of the CardNo in this table must be detected and restored.

NCalculations Table: Used to maintain the checksum values for corresponding CardNo values.

Our approach allows users to perform basic data manipulation operations such as Insert, Update, Delete and Select. Two types of triggers are then used, i.e. conventional triggers [9] and block-type triggers [9].

Conventional Triggers

- **Pre-Query Trigger.** Used to delete any data inserted illegally in “DB1” database by unauthorized user, and to check user authorization, see Figure (5).

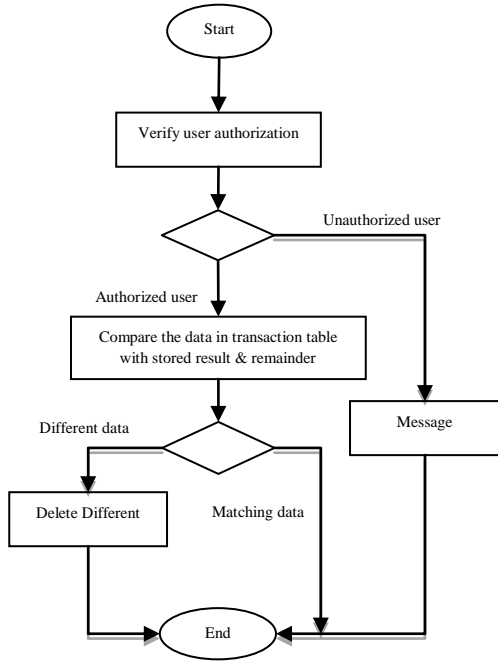


Figure (5)

- **Pre-Insert, Pre-Update & Pre-Delete Triggers.** These triggers check user authorization, and if the user is unauthorised an error message is generated. See Figure (6)

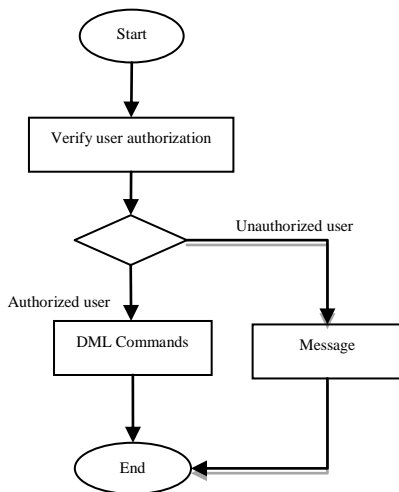


Figure (6)

Block Triggers

- **Pre-Insert Trigger:** This trigger calculates the checksum values i.e., P_i and R_i using equations 1 and 2, and stores the values in the NCalculations table in the “CS1” database. See Figure (7)

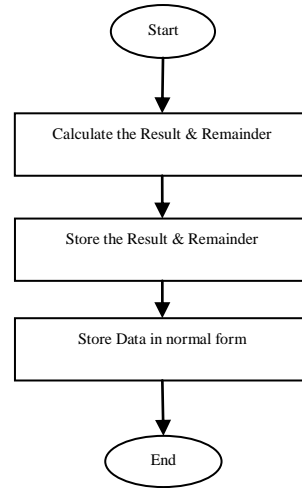


Figure (7)

- **Post-Query Trigger:** This trigger is used after any query to re-compute the stored checksum values using equation (3) to regenerate the original values stored in the “CS1” database. Recomputed values are compared with the values stored in the transactional table in “DB1”. If they are not the same, the recomputed value will be stored in “DB1”. See Figure (8)

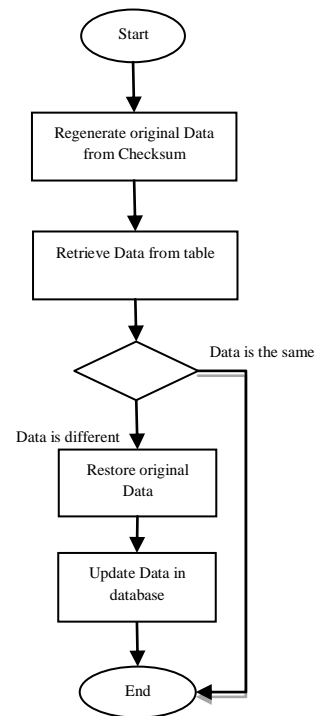


Figure (8)

- **Pre-Update Trigger:** This trigger provides the means to update the values record in “DB1” and “CS1”.

Updates are forced on the Ncalculations table which contains the values P_i and R_i . The update uses the recalculated checksum values i.e. the P_i and R_i . See Figure (9).

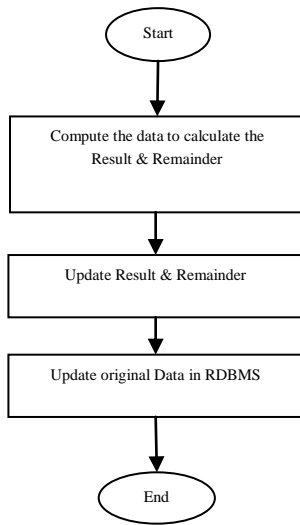


Figure (9)

- **Pre-Delete Trigger.** This trigger deletes any values from the transactional table in the “DB1” database and the related checksum values in the “CS1” database, i.e. delete the P_i and R_i

4.3 Character implementation

This implementation example involves the tables in Figure (10). The **Client Table** contains the client transactions information. Illegal changes to the ClientName in this table must be detected and restored. The **Calculations Table** stores the checksum values of P_i and R_i , i.e. the result and remainder.

The implementation must provide users with a means of performing the basic DML operations Insert, Update, Delete and Select. Two types of triggers are used, i.e. f Conventional triggers [9] and block triggers [9] of the same kind used the implementation in Figures 5 and 6.

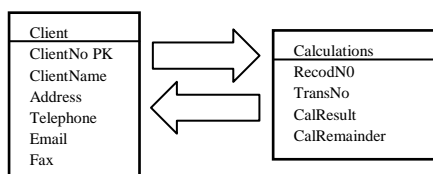


Figure (10)

Block Triggers

- **Pre-Insert Trigger:** This trigger converts the text ASCII codes and calculates the checksum values P_i and R_i which are then stored in the Calculations table in the “CS1” database. See Figure (11) below.

- Pre-Update Trigger:

This trigger provides the means to update the values in the “DB1” database as well as the checksum values in the “CS1” database. Updates are forced on the Calculations table containing the checksum values P_i and R_i . See Figure (12). This update uses the recalculated checksum values for P_i and R_i .

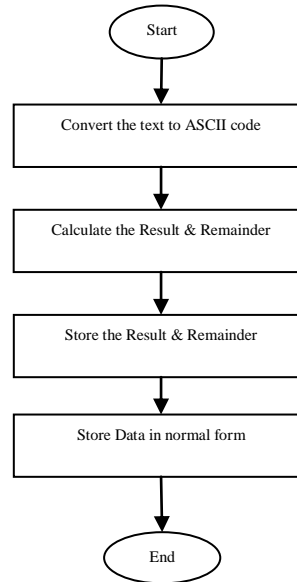


Figure (11)

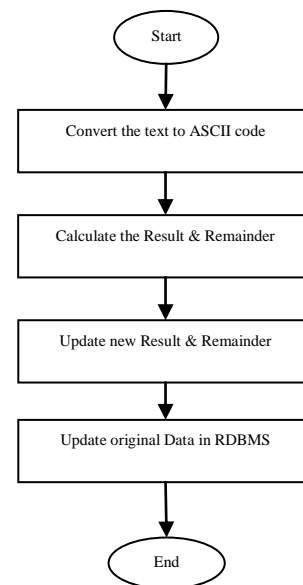


Figure (12)

- **Post-Query Trigger:** This trigger is used after all queries and re-computes the stored checksum values using equation (3), i.e. it re-computes the ASCII values and then converts the ASCII codes to a string. The checksum values in “CS1” are then compared with the data stored in the transactional table in “DB1” and, if they are unequal, the

correct value is automatically inserted into “DB1”. See Figure (13).

Using our approach, conventional user authorization is maintained via a Username & Password combination which must be provided when connecting to a database. In addition *Roles Assignment* [9] ensures that authorized users are granted the required role and that their transactions are thus valid. In our examples, users were assigned the role “CHECKSUM” in order that they are able to make legal changes.

5.0 Test and Results

As a simple test to determine the correctness of the proposed approach, 11 random records have been entered into a “DB1” database. The test data was then updated by illegal users using direct access to the “DB1”.database through SQLPLUS where all ClientName and CardNo attributes were updated and committed as shown below.

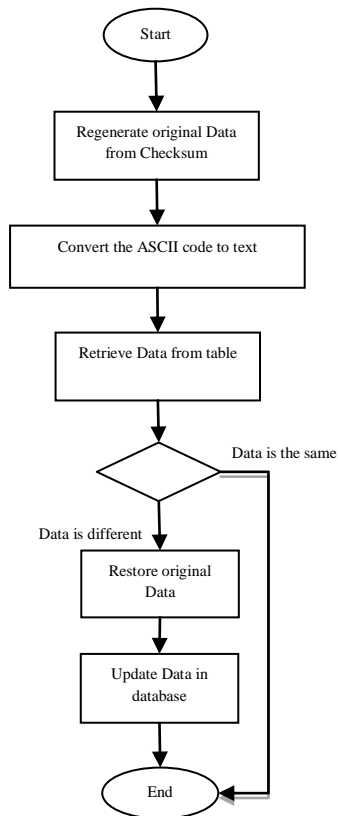


Figure (13)

5.1 Character Test

The following SQLPLUS command was used to check the client name from the table called client.

```
SQL> select clientname from client;
```

The returned result as displayed below:-

```
CLIENTNAME
```

```
-----
Allen Marlard
Ward Jones
Martin King
Adams Blake
Martin Jones
Scott Clark
King
Turner King
Adams
James Miller
Miller King
```

11 rows selected.

Assuming the client names in the table client are to be changed illegally by directly connecting to the SQL to “DB1”database, the illegal update to be executed and committed requires the following commands:

```
SQL>update client set clientname =' Oliver';
11 rows updated.
```

```
SQL>commit;
```

commit complete.

If we recheck the constants in the table client in the “DB1”database, all client names have been changed to OLIVER as shown below.

```
SQL>select clientname from client;
```

```
CLIENTNAME
```

```
-----
Oliver
Oliver
Oliver
Oliver
Oliver
Oliver
Oliver
Oliver
Oliver
Oliver
Oliver
```

11 rows selected.

It should be mentioned here that the critical data in the “DB1” database at each query by the proposed application is checked by re-computing the checksum values and comparing them with the data stored in the “DB1”database. If the data values are equal the content is considered to be OK, however if they differ the original values are restored using the computed checksum values. Assuming that a legitimate user queries the “DB1” database using our application, and if at the same time checks are made on the

content of the table *client* it can be seen that the original data has been restored.

We select the ClientName again from SQLPLUS as below.

```
SQL> select clientname from client;
```

```
CLIENTNAME
```

```
-----  
Allen Marlard  
Ward Jones  
Martin King  
Adams Blake  
Martin Jones  
Scott Clark  
King  
Turner King  
Adams  
James Miller  
Miller King
```

11 rows selected.

This test determines if an illegal user inserts a record into the “DB1” database, and, if so, the illegal insertion is removed from the “DB1” database as follows.

First the record is inserted and committed:-

```
SQL> insert into client values (100,'CATHIE','88 Park  
Road',  
'0161383621','Cathie@hotmail.com','0161383620',1);
```

1 row created.

```
SQL> commit;
```

Commit complete.

A recheck of the table *client* then indicates that *Cathie* has been added in the as shown below.

```
SQL> select clientname from client;
```

```
CLIENTNAME
```

```
-----  
Allen Marlard  
Ward Jones  
Martin King  
Adams Blake  
Martin Jones  
Scott Clark  
King  
Turner King  
Adams  
James Miller
```

Miller King

Cathie

12 rows selected.

Queries on the “DB1” database using our application forms involve a check on the content of the table *client*, and it can be seen that the original data has been restored and the illegal record has been removed, i.e. selecting the *clients* as shown below:

```
QL> select clientname from client;
```

```
CLIENTNAME
```

```
-----  
Allen Marlard  
Ward Jones  
Martin King  
Adams Blake  
Martin Jones  
Scott Clark  
King  
Turner King  
Adams  
James Miller  
Miller King
```

11 rows selected.

5.2 Number Test

The same type of test was performed on the CardDetails table. The results were the same as the character test. In the first test, the CardNo was updated by an illegal user and it successfully restored, in the second test a new record was illegally inserted, then detected and removed.

The two tests have been carried out on an implementation of out model. In both tests the system suffered from illegal access via direct access to the “DB1” database where several records were updated and/or new records inserted in the Client and CardDetails tables.

The test results indicate that the model is working, and the performance of the system was monitored during tests and was not adversely affected by the overhead of computing and using checksum values.

Table (4), shows the actual processing times, for the update comparison operation on the test database, the test was carried out using different database sizes. As it can be seen that the data integrity recalculation process carried out affects the performance of the system, however this effect is within the acceptable time margins and would not affect user during a normal run of the system as the delay is negligible.

No. of Records	Scenario1 regular DB Milliseconds	Scenario2 DB with Checksum Milliseconds	Checksum Overhead
10,000	31,236	68,406	2.19
20,000	62,187	137,813	2.22
30,000	94,000	203,765	2.17
40,000	127,406	271,765	2.13
50,000	158,094	340,219	2.15
Average			2.17

Table (4) Test Results

Figure (14) shows the actual test data representation of both scenarios the first being without checksum parameters and the second scenario with the checksum parameters.

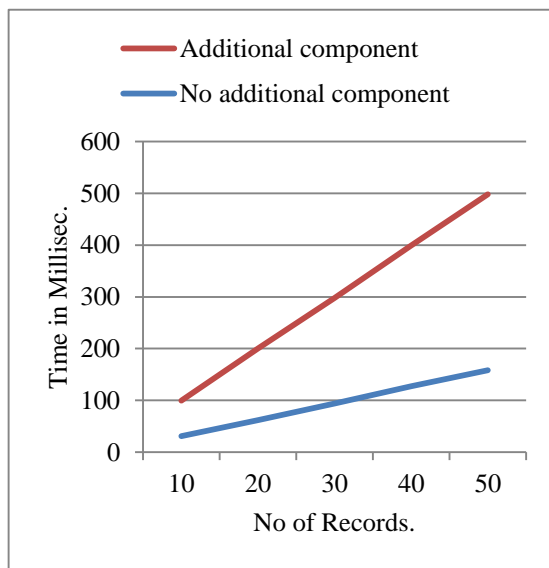


Fig (14) Oracle 10G Checksum Update overhead

6. Conclusions

We have presented an approach that can be used for all types of data; this approach is based on using a checksum validation algorithms applied only to critical data such as credit card numbers. This approach protects the data from any illegal changes and guarantees data integrity. In this case the checksum values can be stored in the same database as the regular data or it can be stored in a separate database i.e. "CS1" database, this will increase the level of security. It was also shown that the algorithm provide the necessary data integrity needed for any system regardless of it size and or data type. The proposed data checksum mechanism has an advantage that it can detect whether data has been modified, and in this case the algorithm compute the original data, this will provide a data integrity mechanism to protect the data.

Future work can be implemented based on this algorithm were the integrity of encrypted data can be maintained, in this case the checksum values can be calculated for the encrypted data, the checksum values are then stored to be used at any time to reconfirm the encrypted data validity. This will protect encrypted data from any illegal changes that might take place by unauthorized users. Therefore this algorithm is capable of protecting any data type from illegal access and or changes, especially changes such as replacing data by similar type but different in value into the database.

References

1. Jingmin He and Ming Wang. Cryptography and Relational Database Management Systems. Proceedings of the International Database Engineering and Application Symposium (IDEAS'01), 2001, IEEE, Computer Society.
2. Silberschatz, Korth, Sudarshan. Database System Concepts, 4th edition.
3. Dorothy E. Denning, Cryptographic Checksums For Multilevel Database Security
4. Andrew S. Tanenbaum, Computer Networks, 4th edition. Prentice Hall. 3rd chapter.
5. Gopalan Sivathnu, Charles P. Wright, and Erez Zadok, "Enhancing File System Integrity Through Checksum", Technical Report FSL-04-04, Stony Brook University.
6. Darren R. Crocker, 2004, Figurehting Spam on Lotus Notes/Domino using no.Spam.domino, Block Checksum detection, Whitepaper.
7. Darrin Chapman and Toby Creek. "Oracle and Open Systems SnapVault Backup and Restore for UNIX Platform". Network Appliance Inc. TR 3377, 2005.
8. Maitrayi Sabaratnam, "Cost of Ensuring Safety in Distributed Database Management Systems". Norwegian University of Science & Technology.
9. Oracle8i Application Developer's Guide - Fundamentals. <http://www.csee.umbc.edu/help/oracle8/server.815/a68003/toc.htm>
10. Oracle Database Security Checklist, An Oracle White Paper January 2006
11. The Database Holds Your Core Assets—Protect It First, <http://www.devx.com/dbzone/Article/11961/1763>
- 12 Database views, http://publib.boulder.ibm.com/infocenter/wbcr/v251/index.jsp?topic=/com.ibm.wbcr.doc/apidoc/i_bcr_r_apidatawbc.html

Coupling GIS-based MCA and AHP techniques for Hospital Site Selection

Aseel Kmail

Jacqueline Jubran

Walid Sabbah

Geographic Information System Department, AAUJ

Jenin, Palestine

aseel.kmail@aauj.edu

jacqueline.jubran@aauj.edu

walid.sabbah@aauj.edu

Abstract

Recently, the population of Jenin city is increasing rapidly and this amplifies the need for more infrastructural objects such as hospitals. Hospitals are considered among the most important infrastructural constructions in cities as they provide health care services. However, current hospitals and medical resources are limited and randomly allocated in Jenin city. Accordingly, in this paper we propose a suitability model that employs Geographic Information System (GIS) based Multi-Criteria Analysis (MCA) with Analytical Hierarchy Process (AHP) to identify suitable locations for building a new hospital in Jenin city. An experimental instantiation of the proposed model is instantiated and the produced results show that the majority of suitable areas are located in Northeast of Jenin. This is mainly because Northeast of Jenin is away from industrial areas and dumping sites.

Keywords: Hospital site selection, GIS, MCA, AHP.

I. Introduction

Recently, the number of people shifting from rural to urban areas is increasing rapidly. This rapid increase of urban population creates various social, economic and environmental changes such as unplanned sprawl, inadequate housing facilities, traffic congestion, insufficient drainage, and lack of health facilities [1]. It is very important to provide all facilities and infrastructural constructions in urbanized areas to overcome the rapid urban growth. Therefore, it becomes the government's responsibility to provide the required resources and facilities for urban areas on proper locations.

Hospitals are among the most important facilities that have a vital role in providing health care services. Identifying the best locations for new hospitals is an important issue due to the fact that selecting suitable locations will help the government to optimize the allocation of medical resources, simplify social contradictions and control the health care development in rural and urban areas. On the other hand, appropriate hospital site

selection will help people reach hospitals easily, reduce the time of rescue and improve the quality of life [2].

Current hospitals in Jenin city are randomly distributed and arbitrarily allocated due to the unreasonable distribution by the government. For example, inside the city center of Jenin, hospitals are saturated. With the growth and extension of Jenin, the population increases rapidly and spreads into areas outside the city center and the contradiction between supply and demand for hospitals is becoming severe. Moreover, there is a persistent need to build quality hospitals that provide professional health care services due to the limited high quality medical buildings.

Several studies employed GIS techniques and products to address the problem of identifying the best locations for building hospitals and planning health services [3, 4]. Most of these studies take into account several parameters to allocate suitable sites for building hospitals such as existing hospitals, population, economical factors, pollution, and other laws and regulations.

In this paper, we aim to identify the most suitable areas for building a new hospital in Jenin city. In order to achieve this goal, we will exploit GIS products and methods with MCA in addition to AHP. By this, we mean that the study will take into account many factors such as existing hospitals, proximity to main roads, and distance to polluted and industrial areas. After that, we will assign them different weights (according to their importance) based on AHP.

The main contributions of our work are summarized as follows:

- Employing GIS-based MCA in order to identify the best locations for building a new hospital in Jenin city.

- Exploiting AHP to assign weights and scores for the identified criteria (i.e. factors) in order to select the best location for the new hospital.

The rest of this paper is organized as follows. Section 2 presents the related work. A general overview of the study area is presented in section 3. Section 4 presents the general architecture of the proposed model and the implementation details of the suitability model. Experimental validation and evaluation of the proposed model is presented in Section 5. In Section 6, we draw the conclusions and outline future work.

II. Related work

We will clarify our contributions in the following paragraphs by offsetting them with prior related work. Several studies have employed GIS techniques and methods in health services and for planning public health [3, 4 and 5]. For example, the authors in [6] combined GIS with Location Based Services (LBS) in order to settle the affairs of emergency medical incidents. On the other hand, other authors employed GIS techniques and methods in selecting the best site for building health care facilities. In order to build constructions that provide health care facilities, various parameters (i.e. factors) can be considered to identify the most suitable sites such as existing health care facilities, population, economic factors and pollution. These parameters can be classified, analyzed and integrated together in different methods. For example, MCA is used to identify factors that affect building new health care objects in [7]. While in [8], the researchers employed both GIS and Analytical Hierarchy Process (AHP) to determine the parameters that affect the physical accessibility of neurosurgical emergency hospitals in Sapporo city. At the same time, the authors of [9] exploited AHP to evaluate the appropriateness of the location selected for Taiwanese hospital.

Although AHP allows multi-criteria decision-making, it suffers from the fact that there are hidden assumptions like consistency. Besides, it is difficult to use when there is large number of criteria. To overcome these problems, Fuzzy Analytical Hierarchy Process (FAHP) is used later in for hospital site selection [10].

In our proposed work, we aim to employ MCA based on GIS methods and techniques to identify the best site for building a new hospital in Jenin city. Besides, we will exploit AHP to assign weights for the factors that affect the new hospital site selection. According to the produced results, we can prove that GIS-based methods and tools play a vital role in making effective decisions in the health field.

III. Study Area

The study area is Jenin governorate. It is located in the north of West Bank as shown in Figure 1. In 2016, the city had a population of 318,958 according to the census by the Palestinian Central Bureau of Statistics [11]. It is located about 43 Kms north of Nablus, and it is about (100-250 m) above sea level. The name of Jenin was derived from Ein Ganim meaning “the spring of Ganim” and referring to the region’s plentiful spring.

Jenin is under the administration of the Palestinian Authority. Today, Jenin is built on the slopes of a hill and surrounded with different types of trees such as carob, fig, and palm trees. It is distinguished by its agriculture, producing various types of crops.

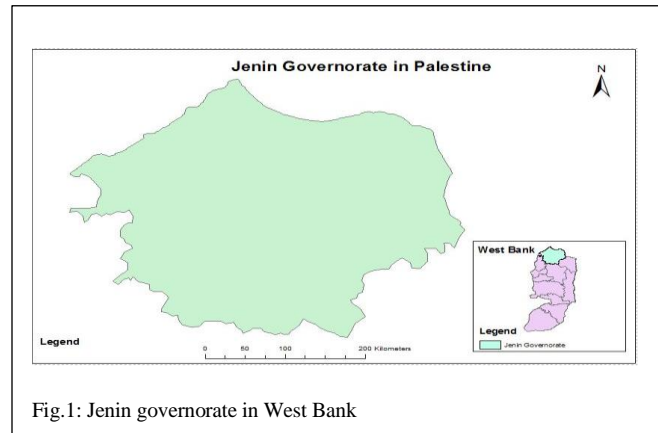


Fig.1: Jenin governorate in West Bank

Jenin governorate has 82 localities and one camp, and we divided the study area into three main regions: Jenin city, Jenin camp and villages that belong to Jenin governorate.

During our work, we focus on Jenin city that has three main hospitals. The details of these hospitals are given in Table 1, and their locations are illustrated in Figure 2.

Table 1: Existing hospitals in Jenin city.

Name	Specialization	Administrated by	No. of beds
Al-Razi Hospital	General	Private sector	60
Al-Amal Hospital	General	Private sector	20
Jenin Government al Hospital	General	Government	120

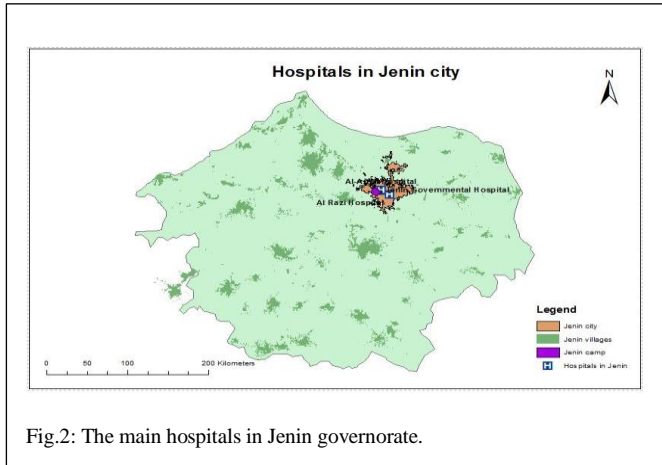


Fig.2: The main hospitals in Jenin governorate.

IV. Data and Methodology

In this section, we present the methodology used in our proposed model in order to identify the optimal site for building a new hospital in Jenin city. Figure 3 depicts the steps of our proposed model.

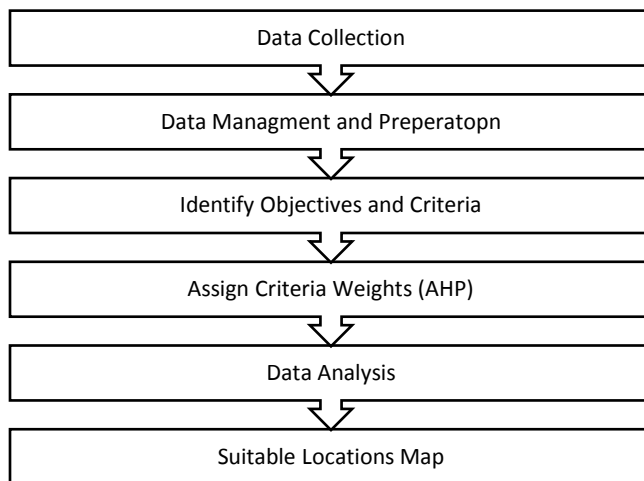


Fig.3: A flow chart depicts the methodology used in order to identify suitable locations for building a new hospital.

As shown in figure 3, the proposed model starts with collecting data from various resources to represent the aspects of the study area. Some of those data are collected from online resources such as GeoMOLG [12] and others are manually digitized with a suitable scale. After that, the collected data are managed and prepared for use. By this, we mean that the data are organized and stored efficiently for further analysis. Then, analysis objectives and criteria are identified to use them in further steps. Once analysis criteria are identified, they are assigned weights based on AHP. Those weights are used to indicate the importance of each criterion. After that, during the data analysis step, various GIS tools and methods

are employed in order to produce the set of suitable locations for building a new hospital.

A. Data Collection

In order to find the best location for building a new hospital, we need to collect data in the format of vector shape files. These shapefiles are collected from GeoMOLG [12] and some of them created by digitizing maps obtained from Jenin municipality. Once the vector data are collected, they are converted later during the analysis step into raster format.

B. Data Management and Preperation

During this step, the collected data are prepared to be used in the analysis process. Data are often collected with missing values and errors, so we need to correct these errors and organize the data in datasets and geodatabases. The process of correcting the data and integrating them into feature data sets constitute a vital role in this step.

Additionally, it is important in this step to answer some questions about the collected data such as:

- What is the data format?
- At what scale it was collected?
- Are the data projected?
- Does the data have all the needed attributes?
- Does the data have constraints and the features geometry support the analysis process?

C. Identify Objectives and Criteria

In our proposed model, we aim to select the optimal site for building a new hospital. Various factors have been involved in the selection process including the following:

- 1) *Land use.*
- 2) *Distance to existing hospitals.*
- 3) *Intersection with main roads.*
- 4) *Distance to dumping sites.*
- 5) *Distance industrial areas.*
- 6) *Elevation.*

These factors are divided into three main types:

- 1- Technical factors: these factors have a clear impact on the construction process and they include the elevation, the slope, distance to existing hospitals and

the land use of the proposed site. The land use refers to how the land being used by human. While the distance to existing hospitals is how fare the new hospital from other hospitals in the same city.

- 2- Environmental factors: there is a strong relationship between human and the environment. The main environmental concerns that may affect hospital site selection are noise and pollution. And thus, the new hospital should be away from noisy and polluted areas such as industrial areas and dumping sites.
- 3- Socio-economic factors: these factors mainly includes proximity to transportation and main roads.

D. Assign Criteria Weights (AHP)

In this step, we assign weights and scores for the identified factors in the previous step based on AHP. AHP has been widely exploited in health-care and medical related problems. The following steps are used to assign weights for all identified factor based on AHP:

- 1- Layout and expose the overall factors.
- 2- All factors can be compared using pair wise comparisons in order to generate weights for factors through distributing questionnaires on experts. In pair wise comparisons, we decide which factor is more and how much important than another using 1-5 scale as shown in Table 2 [13]. The produced wrights quantify the importance of factors in the analysis and decision making process.
- 3- Check the consistency ratios of all pair-wise comparisons.

In this step, we use the Consistency Index (CI) and Consistency Ratio (CR) formulas to check the consistency as follows:

$$CI = (\lambda_{max} - n)/(n - 1) \quad (1)$$

Where:

n: the number of criterion.

λ_{max} : the biggest eigenvalue of the comparison matrix.

$$CR = CI/RI \quad (2)$$

Where:

RI: a constant corresponding to the mean random consistency index value based on n.

- 4- The relative scores are aggregated using geometric mean method.

Table 2: Pair wise comparison scale

Verbal judgment	Explanation	Number
Extremely Un-Important	A criterion is strongly inferior to another	1/5
Moderately Un-Important	A criterion is slightly inferior to another	1/3
Equally Important	Two factor contribute equally	1

Moderately Important	Judgment slightly favor one criterion over another	3
Extremely Important	Judgment strongly favor one criterion over another	5

E. Data Analysis

In this step, a model is developed in order to identify the optimal location for building a new hospital. In this model, the raw data should have the same spatial reference and they are converted into a raster with the same cell size, making them easier reclassified in Analysis steps.

The data analysis steps and tools are detailed as follows.

- Distance to existing hospitals based on network analysis: as detailed earlier, there are six factors taken into account for building our model. In this step, we derive a series of polygons (service areas) that represent the distance that is required to reach each hospital. As a prerequisite to finding the service areas, we need to construct a network dataset. The results of applying this step are depicted in Figures 4 and 5.

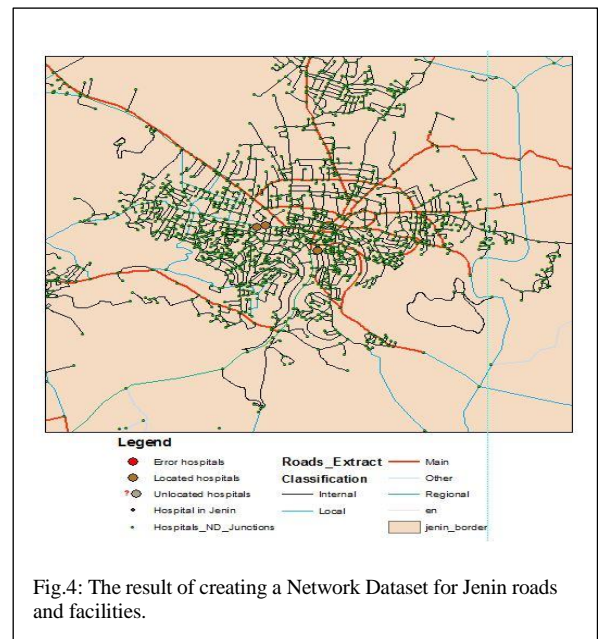


Fig.4: The result of creating a Network Dataset for Jenin roads and facilities.

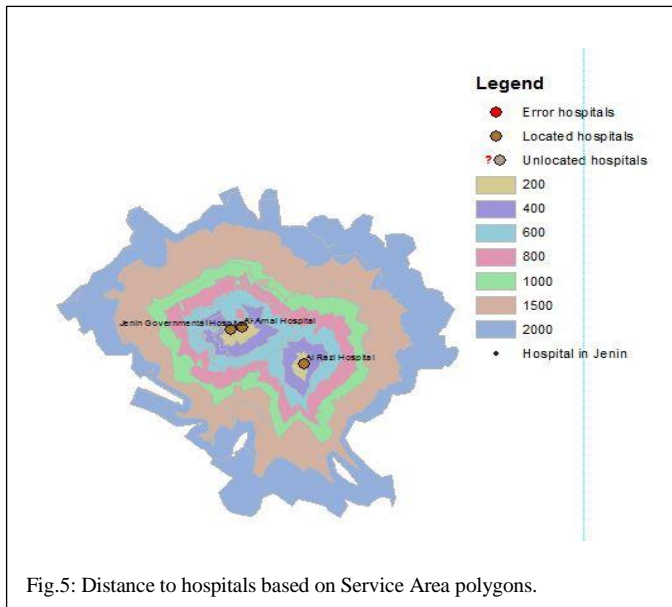


Fig.5: Distance to hospitals based on Service Area polygons.

- Euclidean distance: In this step, we derive the Euclidian distance from facilities (dumping sites and industrial areas) to each pixel in the generated output raster. The formula for finding the Euclidian distance is depicted below:

$$d(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3)$$

Where:

$p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$ are two points in Euclidean n-space.

d: Distance from p to q

The results of applying this step are depicted in Figures 6 and 7.

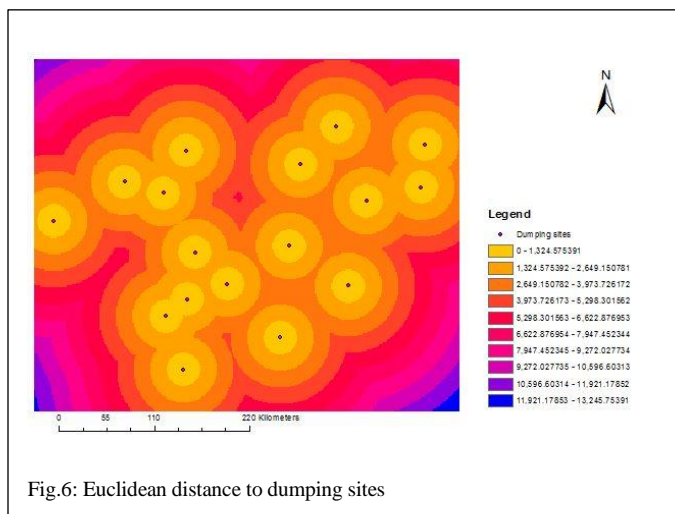


Fig.6: Euclidean distance to dumping sites

- Feature to raster: During this step, we convert the land use feature class (vector data) to a raster that has the same cell size as the derived raster layers from the previous step. Accordingly, we can use all of them for further processing.

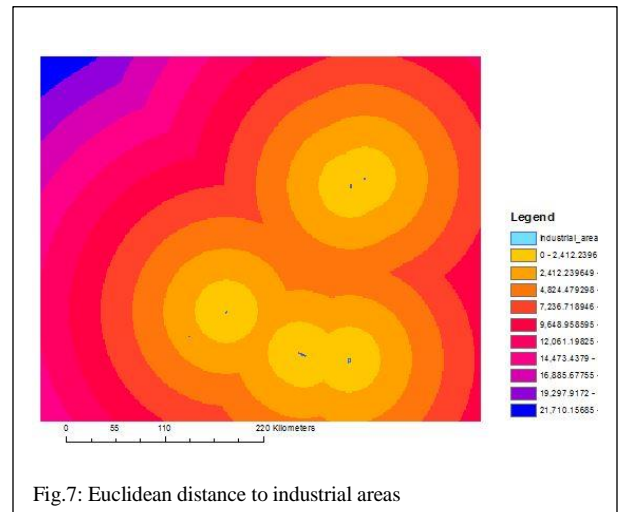


Fig.7: Euclidean distance to industrial areas

- Slope: In order to build a hospital, the land should be relatively flat. Therefore, we consider the slope of the land in our model by deriving the slope of the elevation dataset as shown in Figure 8. By this, we mean that the rate of maximum change in elevations is calculated.
- Reclassification: Each cell in the study area now has a value for the following factors (existing hospitals, dumping sites, industrial areas, land use and elevation). We should combine the derived datasets in order to identify the potential location for building a new hospital in the next step (Weighted overlay). However, we cannot combine them in their current form. For example, there is no meaning to combine cell values that have 15 degrees slope with cell values that have agriculture land use that equals (6). Accordingly, to combine datasets, we need to derive a common measurement scale such as from 1 to 10. This scale identifies how suitable a specific location for building a new hospital. Lower values indicates

locations that are more suitable. The reclassification process is depicted in Figure 9.

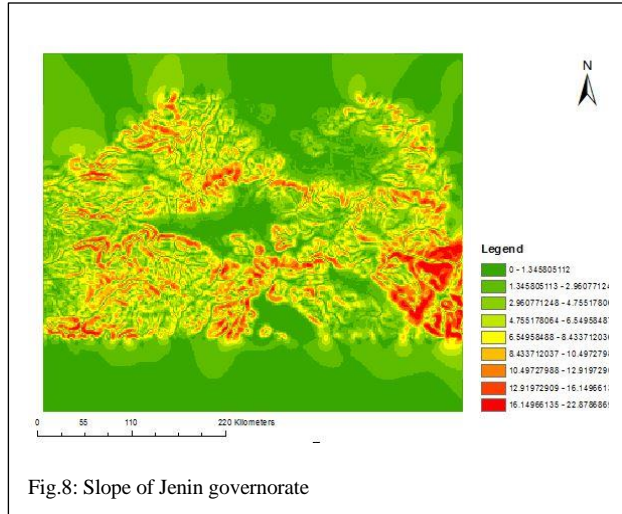


Fig.8: Slope of Jenin governorate

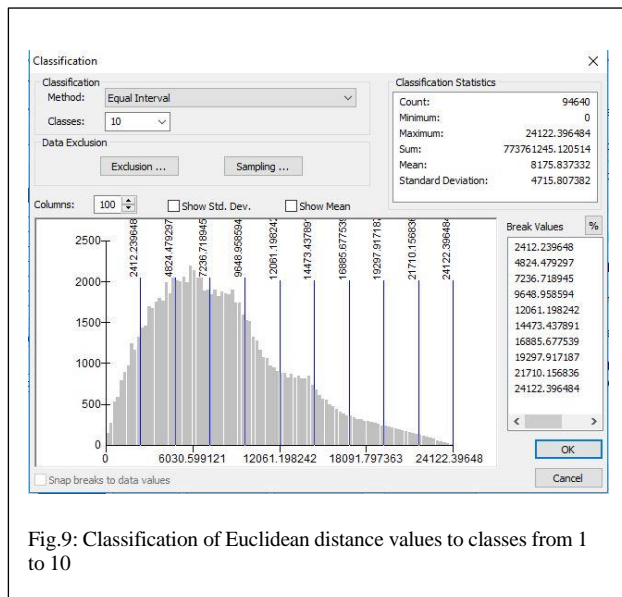


Fig.9: Classification of Euclidean distance values to classes from 1 to 10

As shown in figure 9, we classify the produced distance values from the Euclidean distance process into 10 classes by dividing the produced ranges into equal intervals.

- **Weighted overlay:** Using this technique, we weight the values of each dataset by assigning each a percentage of influence. The higher the percentage, the more influence an input has in the suitability model. Some input values will be restricted. For example, areas that belong to “C” administrative division in West Bank are restricted as shown in Figure 10. The result of this step indicates how suitable each location for building a new hospital.

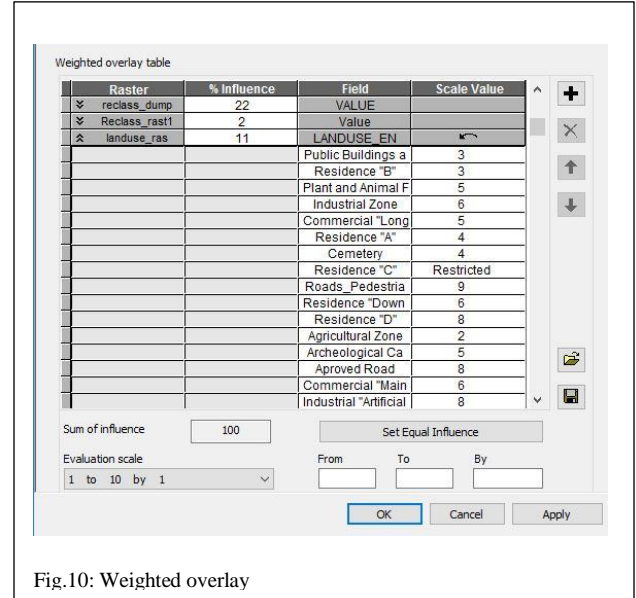


Fig.10: Weighted overlay

- **Majority filter:** The size of the suitable area is an important criterion in identifying the optimal site for building a new hospital. Thus, we use this tool to ensure that the number of neighboring cells of a similar value must be large enough to build a new hospital.
- **Condition:** Pixels with values less than 3 indicates suitable locations. Thus, the condition tool used to identify those locations.

V. Experimental Results and Discussion

This section describes the experiments carried out to evaluate our proposed suitability model. A prototype of the proposed model is implemented and experiments are conducted using a PC with dual-core CPU (2.1GHz) and (8 GB) RAM. The used operating system is Windows 10.

- Comparing the Produced Results When Changing the Weights Assigned for all Input Data Sets

In this section, we compare between the produced results by the proposed model when we change the weights assigned for the input data sets according to the following table.

Table 3: The weights assigned for input datasets in different experiments.

Input data set	Random weights in EXP.1	Random weights in EXP.2	AHP weights in EXP.3
Land use.	0.09	0.10	0.22
Distance to existing hospitals	0.04	0.20	0.13
Near main roads	0.24	0.04	0.30
Distance to dumping sites	0.13	0.06	0.22
Distance industrial areas	0.30	0.30	0.02
Elevation	0.20	0.30	0.11

As shown in table 3, we assigned different weights for each input data set in different experiments. After running the proposed model with weights from the first experiment (EXP.1) and the second experiment (EXP.2), we get the results shown in Figure 12.

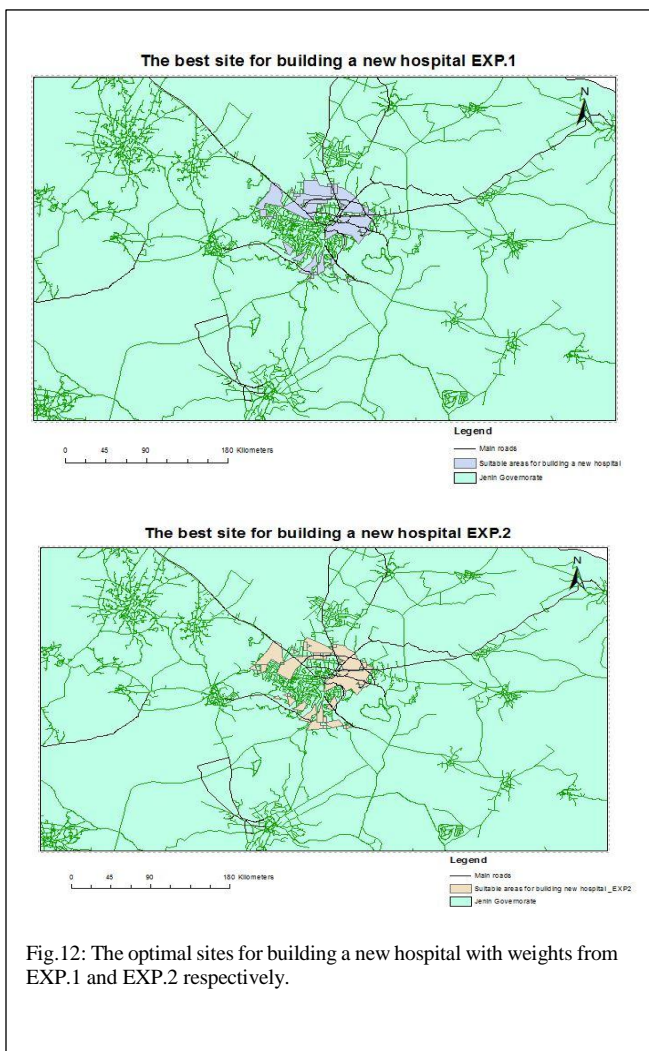


Fig.12: The optimal sites for building a new hospital with weights from EXP.1 and EXP.2 respectively.

As we can see in Figure 12, the produced results have various contiguous alternatives, which may be confusing for the decision maker. This is because the weights are randomly assigned for the input datasets without extensive care or research.

However, we were able to achieve promising results for decision makers when using the weights produced from applying AHP as shown in Figure 13.

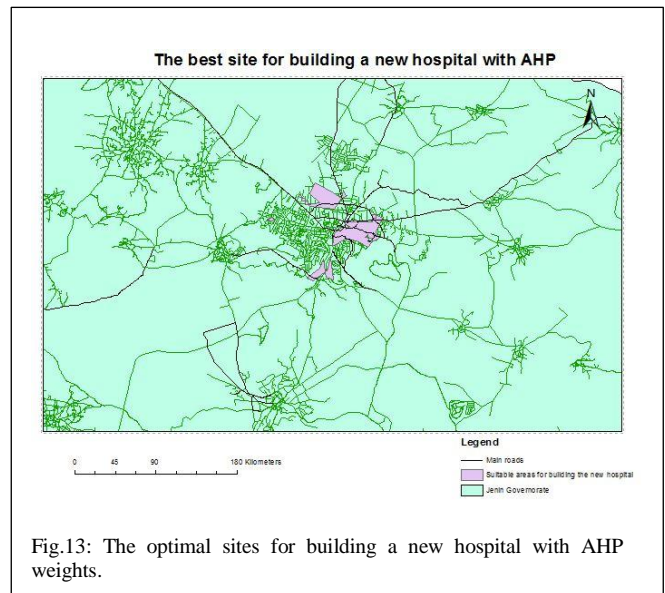


Fig.13: The optimal sites for building a new hospital with AHP weights.

VI. Conclusion and future work

In this paper, we built a suitability model for selecting the optimal site for building a new hospital based on coupling GIS-based MCA and AHP. GIS tool and techniques are employed to analyze the list of identified criteria in hospital site selection. The analysis process incorporates assigning weights for the identified criteria based on AHP. And at the end of the analysis process, the optimal site for building a new hospital is identified. The results showed that assigning weights based on AHP is better than assigning weights randomly for the set of identified criteria.

References

- [1] Liu, Y., 1998. Visualization the urban development of Sydney (1971–1996) in GIS. In: Proceedings of 10th colloquium of the spatial information research centre, University of Otago, New Zealand, 16–19 November.
- [2] Lina Zhou, Jie Wu. GIS-Based Multi-Criteria Analysis for Hospital Site Selection in Haidian District of Beijing. 2012
- [3] Hare, T.S., Barcus, H.R., 2007. Geographical accessibility and Kentucky's heartrelated hospital services. *Applied Geography* 27, 181–205.
- [4] Astell-Burt, T., Flowerdew, R., Boyle, P.J., Dillon, J.F., 2011. Does geographic access to primary healthcare influence the detection of hepatitis C? *Social Science & Medicine* 72, 1472–1481.
- [5] Gordon, A., Womersley, J., 1997. The use of mapping in public health and planning health services. *Journal of Public Health Medicine* 19(2), 139–147.
- [6] Maglogiannis, I., Hadjiefthymiades, S., 2007. EmerLoc: location-based services for emergency medical incidents. *International Journal of Medical Informatics* 76, 747–759.
- [7] Nobre, F.F., Trotta, L.T.F., Gomes, L.F.A.M., 1999. Multi-criteria decision making: an approach to setting priorities in health care. *Symposium on statistical bases for public health decision making: from exploration to modeling* 18, 3345–3354.
- [8] Ohta, K., Kobashi, G., Takano, S., Kagaya, S., Yamada, H., Minakami, H., Yamamura, E., 2007. Analysis of the geographical accessibility of neurosurgical emergency hospitals in Sapporo city using GIS and AHP. *International Journal of Geographical Information Science* 21(6), 687–698.
- [9] Cheng-Ru, W., Chin-Tsai, L., Huang-Chu, C., 2007. Optimal selection of location for Taiwanese hospitals to ensure a competitive advantage by using the analytic hierarchy process and sensitivity analysis. *Building and Environment* 42, 1431–1444.
- [10] Vahidnia MH, Alesheikh AA, Alimohammadi A. Hospital site selection using fuzzy AHP and its derivatives. *Environ Manage.* 2009 Jul;90(10):3048-56.
- [11] Palestinian Central Bureau of statistics.
- [12] http://www.pcbs.gov.ps/Portals/_Rainbow/Documents/jenna.htm.

Big and Connected Data Analysis with Graph and Relational Databases Using Collaborative Filtering Technique

Joseph George Davis

*Department of Computer Science
Kwame Nkrumah University of Science and
Technology
Kumasi, Ghana
spomegasys@gmail.com*

Joseph K. Panford

*Senior Lecturer, Department of Computer Science
Kwame Nkrumah University of Science and
Technology
Kumasi, Ghana
jpanford@yahoo.com*

James B. Hayfron -Acquah

*Senior Lecturer, Department of Computer Science
Kwame Nkrumah University of Science and Technology
Kumasi, Ghana
jbha@yahoo.com*

Abstract- Over the years systems or applications in existence have been able to work seamlessly with relational databases. Applications such as point of sale, hospital management systems, Human resource applications, payroll systems, banking with relational databases banking systems just to mention a few are powered by relational databases with minimal to entirely no issues because most the relationships amongst entities in these systems are mostly not complicated or highly connected. In the same vein relational database systems have been able to handle large amounts of data and transactions emanating from everyday operations of these systems. In summary the relational database was very effective in dealing with the problems applications available were solving. In these modern times the influx of social media platforms, map and navigations systems, geospatial information systems, recommendation engines ,referral systems and the likes have turned the tide for systems and the databases behind them to support, manage and model mostly semi/unstructured, connected and their

complex associations amongst data elements. This study looked at such data with the concept of recommendations to test whether relational databases were still performing well with the current trends on connected data or the NoSQL paradigm making inroads in the technology space had a point when they say “they are the database paradigm for the future”. An experiment was therefore performed with a relational and graph database to ascertain this analogy.

Key performance indices (KPIs) such as runtime, storage, schema flexibility, query complexity and general operations of each database paradigm were tested with the concept of making recommendations to a number of people in the database based on how their associated(friends) with each other. Each paradigm was put on similar test against the above KPI's and the graph database seemed to have an urge over the relational database in the result and analysis of the figures that were obtained from the experiment.

Keywords: Key Performance Index, Graph Databases, NoSQL, Relational Databases, DBMS

I. INTRODUCTION

The battle between relational [9][11] and non-relational databases began a while ago but technological and software application trends happening now and also what is perceived for the future has intensified this battle. Applications now are thriving more on information which relates to how stored data is connected and associated with each other for further analysis, decision making and recommendations. A direct example is the social media platforms available nowadays hence database management systems behind these systems have to be in the position to meet these ever changing trends and requirements [6]. Each database paradigm (relation and non-relational) has dealt with this issue in its own peculiar way which may or may not have worked in certain circumstances. So the question still remains which of these database systems better handles connected, associated data better for applications that need to interpret this data for analysis and decision making to work in a fast and efficient way, using less processing power, memory, storage space and with less complex queries[1]. The problem most often stems from when the connected and associated data being used by these applications (social media platforms, map and navigation systems, geospatial information systems, recommendation engines (mostly behind e-commerce platforms), referral systems, Internet of things starts growing and complex questions and analysis are requested from the applications to drive short term to long term strategic decision making for organizations. . It is in such situations that the ACID [9] (atomicity, consistency, isolation and durability) features of a database are really tested. So the question still remains which type of database systems (relational or non-relational) is for the future in terms big and connected data which seems to be the bane of applications coming up each single day.

This study therefore aims to test how each of the database paradigms will behave in terms of the parameters noted below.

- General operational differences
- Storage analysis on both DBMS
- Runtime analysis of both DBMS interacting with associative data
- Query complexity analysis
- Handling of schemas for connected data sets.

II. JUSTIFICATION OF THE STUDY

The purpose of this study is to test and profile a relational [9][11] and graph database[16] working behind a recommender application which works on analyzing the associations between data constituents for decision making and strategic planning. In other words, scenarios will be setup where by people having relations with others in the database are

suggested products that friends have bought or liked using relational and graph queries[1][11]. In a nutshell future applications which will be depending mostly on strong analysis of connected data will be built knowing which database paradigm will support and answer the complex questions that will be posed to such applications.

On a higher level application throughput will be tested as query time for complex questions/situations which are solved by such applications tracing and analyzing patterns in user associations in data presented to predict future behavior or preferences. Also amount of physical storage to be used by DBMS(relational and graph database) as data grows in certain proportions in such situations will also be a question to be answered as well as how fast the application will run when data increases. In other words which database paradigm will survive with large amount of data with complex associations? Moreover, a test will also be made if the combination of the two databases backing the same application may or may not help solve complex situations better. This will base on leveraging the strengths of each database uncovered in the course of the study to mitigate the shortfalls of the each other.

III. LITERATURE REVIEW

Thought relational database concept proposed by E.F. Codd [9] has thrived over years and has stood the test of time giving it a merited status as a matured database system, [6] with the recent introduction of systems like the World Wide Web

[15], social media, social networking applications [7], internet of things and the likes have put this maturity to test, raising questions of in database domain in relation to flexibility when it comes to data sets used in these modern applications[8] . This trend has caused the emergence of new database paradigms like NoSQL [14] trying to mitigate the short falls of the relational databases in this regard has a lot variants of databases from columns store to key-value database systems. This study looks at the graph database in relation to the relational database on how they handle related data objects for systems that need it for decision making. Related studies like one done by Emil Eifrem [4][12] looked at comparing graph and relational database in regards to a social example of a 1000 customers/users with an average of 50 friend connections with degrees of connections giving friend-of-friend, friend-of-friend-of-friend down to the 4th hop scenarios. Queries run against each database produced execution time of 2ms for graph and 2000ms for the relational database for a small dataset. The scenario got worse for the relational database since it had to be stopped after a day of running the same query which the graph database run in 2ms.From the study the graph database proved better in that regard.

Looking at a similar research by Sharma M. & Soni P. [1] to analyze the performance of both relational and graph databases with some predefined data processing queries based on a schema to analyze data processing and analyze time with various number of inputs of data objects. The schema used was modeled around a user, friend, movie, actor entities with the queries that need to get all friends of a user, favorite movies of a user's friends and finding lead actors of the movies watched by a user's friends. The study goes on to run these three queries against increasing datasets of 100,200,300 with the relational database having a greater runtime than the graph database at each three instances. B. Shalini & Tyagi C [6] evaluates graph (Neo4j) and relational (MySQL) databases based on evaluating parameters like level of support/maturity, security of the database and database flexibility. In terms of maturity an argument is made for the relational database for providing storage and robust support for commercial applications or products for over a decade as compared to graph databases which came into the technology limelight in the early 2000s so may not be at the same wavelength as relational databases when it comes to production testing over a long period of time. Looking at the security view point the relational database with its built-in multiuser support feature and restrictions is in better shape than the graph database because of its comprehensive support for Access Control List (ACL). Nevertheless the graph database takes lead when it comes to flexibility because of the short fall of relational databases to extend schemas or databases [10]. Relational database also lags when it comes to management of flexible schemas that change over time [13]. As most studies seem to always target just the runtime of queries, this study also looks at that and extends the parameters to add storage analysis, compare schemas and investigate query complexities

IV. METHODOLOGY

For the analysis of connected data one has to know what connected data is. It is mainly data that has individual entities interconnected with each other such that decisions and analysis are made based on the connected relationships between the entities[6][12]. The focus of this study is on such connected data and how it is analyzed by graph and relational databases taking into considerations the size of the data or rows involved from thousands of records, all the way up to a million interconnected records.

This study will be centered around a social relationship [7] amongst friends based on which decisions or recommendations will be made to highlight the base concept of connected data. The recommendation technique used in this study to inform the composition of graph and relational

database queries is the Collaborative Filtering (CF). CF also referred to as social filtering, filters information using recommendations of other people (mostly friends and acquaintances). For example a person who wants to watch movie may ask for recommendations from friends. The recommendations of friends who have similar interests are trusted more than recommendations from others [18]. The data used for the study have been modelled around CF concept to ascertain which of the database paradigms works well in situations where connections among data entities are used to make recommendations. The data that will be looked into can be from a wide variety of products from a hypothetical ecommerce site but it will be towards users purchasing movies, books, games and electronic gadgets. Figure 1 is the underlying general data model.

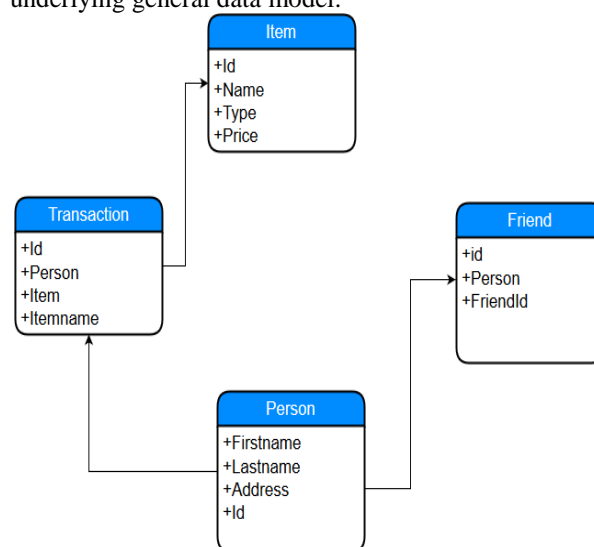


Figure 1: The general data model irrespective the type of database used.

Figure 1 is the underlying general data model which projects four entities (Person, Friend, Item, and Transaction) for the study. This data model portrays people and their friends, how they relate to the items in a hypothetical online store from which recommendations will be made by the system using CF. A person buys or likes an item (s) be it a movie, book, game or electronic gadget. A person also has friend(s) who also has liked or bought an item(s). People can have a lot of friends, buy or like a lot of items so we have large amounts of data with entities associating or relating to each other. The question is how do we leverage on the features of a relational or graph database to find patterns in data that can help recommend or suggest items and even friends to people.

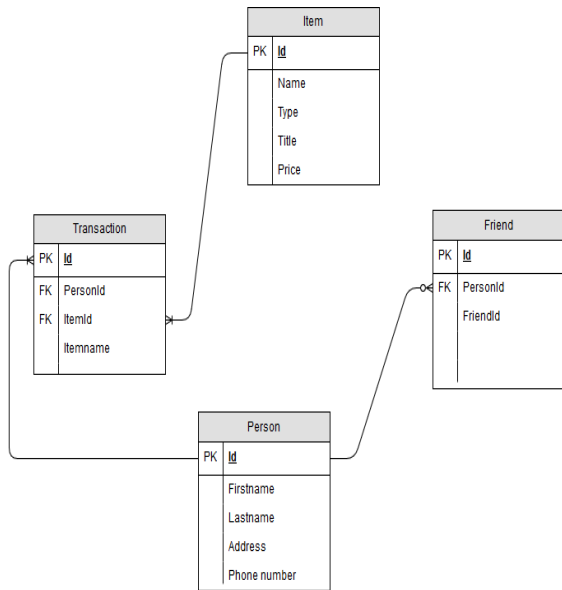


Figure 2: Relational Database Model

Figure 2 shows the relational data model for the system. This shows the relationships the person entity has with the other entities through the foreign key constraint. This further elaborates that a person has friends and can buy or like an item which can be a book, game, movie or a gadget etc. In the relational database paradigm [9] this relationship is shown with a foreign key linking the person entity to the other entities. This is how relationships are portrayed in the relational sense showing connectivity amongst individual entities.

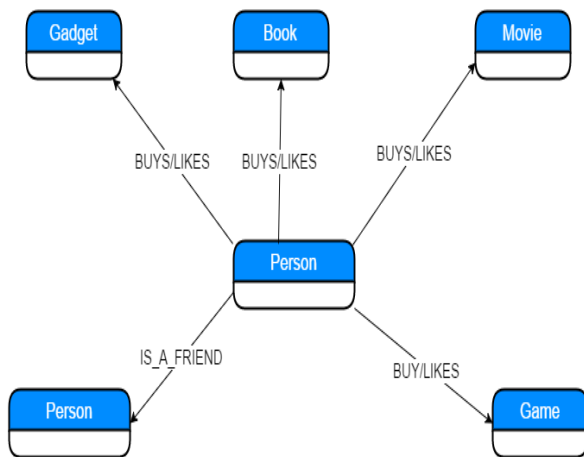


Figure 3: Graph Database Model

The figure 3 above depicts how data modeled in graph [2][5][10] looks like. In graph sense each entity is represented by a node with its properties being the attributes of that node. The relationships between nodes are represented by the labeled vertices between them [16].

This study looks at how these relationships between people, their friends and products can be leveraged

to answer questions like “What are the products a person’s friends are buying” or “What products are a person’s friend of friends buying” which is the basis of recommendation systems to recommend products to people based of their relationship and buying patterns of friends, friends of friends etc.

The system implementation in code was based on the below architecture. The system comprises of a JAVA Enterprise Edition Application riding on a graph (Neo4j) and relational (MySQL).

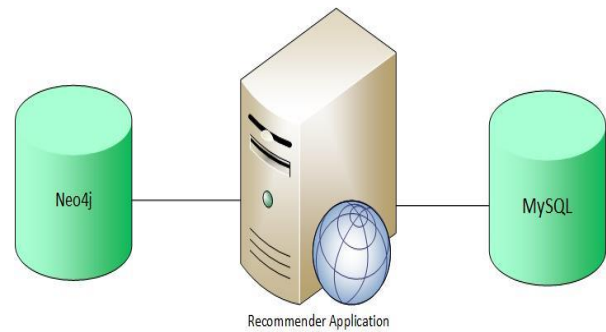


Figure 4: General System Architecture

From figure 4 the system comprises of:

I. Application Server is a JAVAEE application deployed and running on a Glassfish server where the business logic for the application resides. This is where all decisions are made using the information obtained from the attached databases (MySQL and Neo4j).

II. Neo4j [16][13] Graph Database is Graph DBMS that interfaces with the graph implementation side of the Application. Neo4j can be implemented in server mode or embedded mode. For this study/application it is used in embedded mode meaning the management of the database is embedded in code using the Neo4j libraries this means where ever the code is deployed an instance of the Neo4j database is created.

III. MySQL is the Relational DBMS that interfaces with the relational implementation side of the Application.

The systems also works with the below functional requirements:

I. This system works with a graph and relational database to help make product suggestions to people based on the pattern of purchases or likes other people and friends have been doing (Collaborative Filtering)[14].

II. Based on the behavior in (I) above each database will be populated and tested with increasing amounts of data (i.e 10000 people, 100000 people or 1000000 people) to answer questions like “what have friends of a person bought or liked so that recommendations can be made for that person.

III. Analysis on the runtime, amount of storage used and complexity of queries from the database systems used from the data and questions posed to the system in (II) above will be looked at.

Any system that the application can be deployed on must satisfy the requirements below

Software:

- JAVA JDK 7 and above
- MySQL RDBMS
- Neo4j is run in embedded mode
- GlassFish version 4 and above

Hardware: System will be tested on a computer with the specification below

- Intel Core i3 processor
- 8GB RAM
- 500GB hard disk space
- OS Linux Ubuntu 16.04

V. RESULTS

Looking at the runtime analysis of the same query run with a start of hundred transactions all the way to a million with the query average runtime calculated in each scenario. The below results were attained for relational & graph databases:

Table 1: showing the runtime against data results for relational database

data	Execution time/secs
100	0.0028
1000	0.016
10000	0.164
100000	1.315
500000	6.244
1000000	12.896

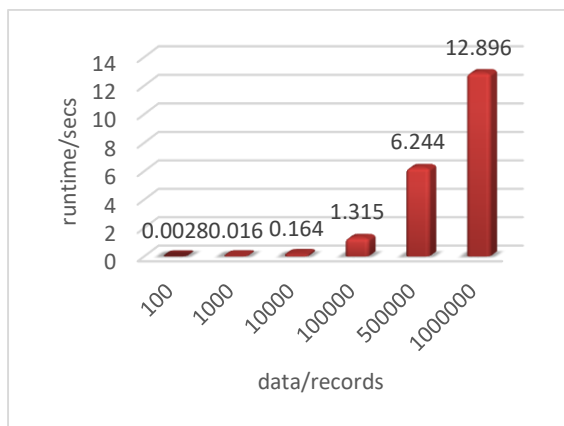


Figure 5. Chart showing amount of data run in query against runtime for a relation database

1

1

Table 3: showing the storage against data results for relational database

data	Storage space/kbs
100	792
1000	840
10000	2744

Table 2: showing runtime against data results for graph

data	Execution time/secs
100	0.002
1000	0.006
10000	0.09
100000	0.219
500000	1.616
1000000	1.728

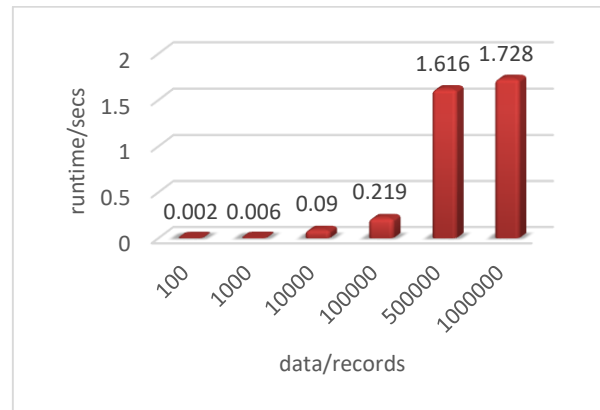


Figure 6: Chart showing amount of data run in query against runtime for a graph database

Figure 7 shows runtime analysis trend of relational and graph databases as data increases.

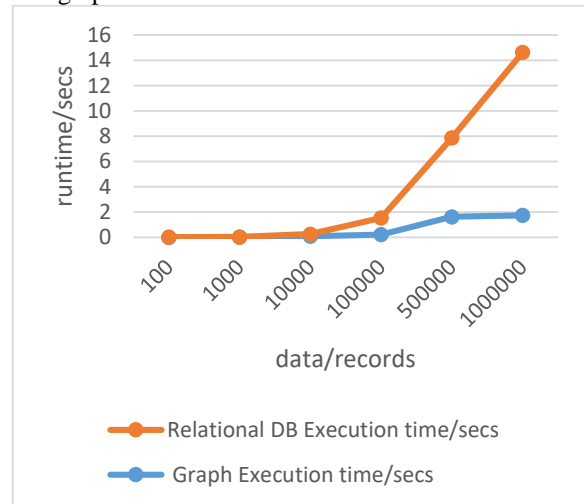


Figure 7: Runtime analysis trend

The same analysis is done with the storage space used by each database shown in figure 10.

100	792
1000	840
10000	2744

100000	14024
500000	37596
1000000	58096

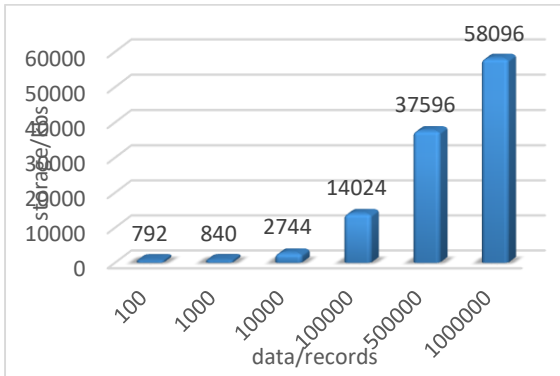


Figure 8. Chart showing database storage size against record size all in kilobytes

Table 4: showing results for graph

data	storage/kbs
100	400
1000	2000
10000	18200
100000	105100
500000	535000
1000000	839000

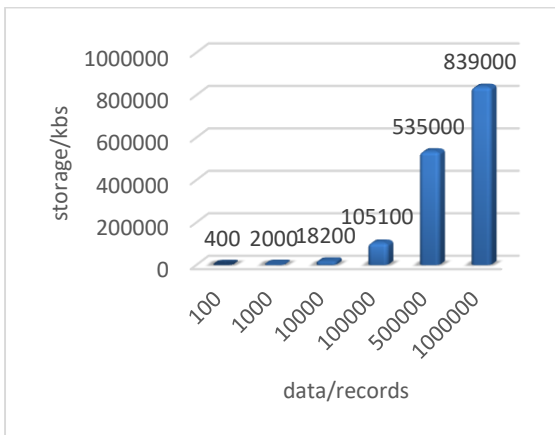


Figure 9: Chart showing database storage size against record size all in kilobytes

Combined Storage Analysis Figure 10 shows the storage analysis trend of relational and graph databases as data increases.

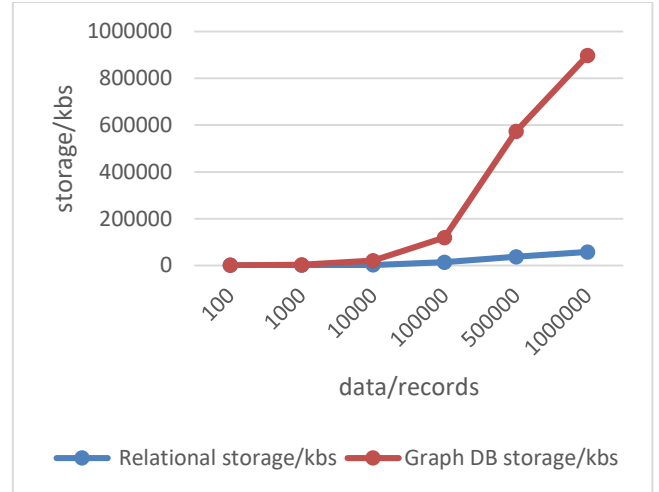


Figure 10: Storage space used by Graph & Relation database

Looking at the trends in figures 7 and 9 it can be seen from the runtime metric trends that as data or number of records increase runtime on each database systems behave differently, for the relational database runtime starts increasing exponentially as the data size reaches 100000 and over hitting as much as 12secs with a million records which is worrying in terms of system performance with large datasets. The graph databases on the other hand maintains a steady increase in runtime as the data sizes reaches 100000 and maintains a runtime of under 2secs even with a million records making it a good option for systems that do large and connected data sets analysis. The story is different when it comes to the amount of space needed by each database to handle large sets of data. It can be seen that the graph database uses a large amount of space as the dataset gets larger while the relational database uses less than 60MB to handle a million records. Even though this development is alarming on the part of the graph database, storage does not seem to be problem for current systems and applications because most of the servers or machines come with not less than 500GB. That notwithstanding it is a big plus for the relational database for its storage optimization. When we assess the complexity of the queries on both sides we get to know why the relational database has a high runtime. From the query complexity analysis it is seen that when it comes to analyzing connected data the relational database often uses joins to associate entities. This presents a complexity of $O(n^2)$ because join queries are executed programmatically in nested loops while graphs queries have a complexity of $O(\log n)$ due to the traversal of the graph data structure. This explains the runtime patterns in figure 7.

VI.CONCLUSION

The analysis of the study looked at the runtime of queries for the two database paradigms followed by analyzing the storage space used as the size of grows. Also, how both relational and graph databases model connected data in the quest to create a database structure that will fulfil the objective at hand.

Based the on the analysis done it can be concluded that when it comes to dealing with big and connected data the graph database has an urge over the relational database. This makes graph database a great options for future applications which will make decisions based on complex data associations and cause to worry on how far the relational paradigm can carry us. Relational databases were conceived to digitize paper forms and automate well-structured business processes, and still have their uses. But RDBMS cannot model or store connected data and its relationships without complexity, which means performance degrades with the increasing number and levels of data relationships and data size. Additionally, new types of data and data relationships require schema redesign that increases time to market.

REFERENCES

- [1] Mukul S. & Pradeep S. (2014), *Quantitative Analysis and Implementation of Relational and Graph Database Technologies*, *International Journal of Modern Computer Science and Applications*, RES Publications, Vol.2, p: 21-25.
- [2] Williams D., Huan . & Wang W, *Graph Database Indexing Using Structured Graph Decomposition*, Department of Computer Science University of North Carolina, Chapel Hill, Department of Electrical Engineering and Computer Science University of Kansas
- [3] Buerli M. (December 2012), *The Current State of Graph Databases*, Department of Computer Science Cal Poly San Luis Obispo
- [4] Josh A. (2013), *Performance of Graph vs. Relational Databases*, viewed January 2016 from:
<https://dzone.com/articles/performance-graph-vs>
- [5] Angles R. *A comparison of current graph database models. In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering Workshops, ICDEW '12, pages 171–177, Washington, DC, USA, 2012. IEEE Computer Society*
- [6] Batra S. & Tyagi C. *Comparative analysis of relational and graph databases. International Journal of Soft Computing*
- [7] Cudr'e-Mauroux P. & Elnikety S. *Graph data management systems for new application domains. In International Conference on Very Large Data Bases (VLDB), 2011.*
- [8] Vicknair C., Macias M., Zhao Z., Nan X., Chen Y., & Wilkins D., *A Comparison of a Graph*
- [9] Codd E.F, *A relational model of data for large shared data banks June 1970*
- [10] Angles R. and Gutierrez C. *Survey of graph database models. ACM Comput. Surv., 40(1):1:1–1:39, February 2008.*
- [11] Connolly M. T. & Begg E. C.(2005),*Database Systems, 4th Edition, Addison Wesley, UK.*
- [12] Robinson I, Webber J. & Eifrem E. (2015), *Graph Databases, O'Reilly, California.*
- [13] Femy P.F, Reshma K.R, Varghese S.M ,*Outcome Analysis using Neo4j Graph Database, International Journal on Cybernetics & Informatics (IJCI) Vol. 5, No. 2, April 2016*
- [14] Jampana A.K, *New Era of NoSQL Databases, International Journal of Emerging Trends in Engineering Research, Vol.3. No.10, [Pages: 362-367(2015) Special Issue of ICACSSE 2015 -Held on October 30, 2015 in St. Ann's College of Engineering & Technology, Chirala, AP, India*
- [15] Berners-Lee, Tim. *HyperText Transfer Protocol Design Issues, 1991. http://www.w3.org*
- [16] Neo4j. <http://www.neo4j.org/>.
- [17] Wikipedia. <http://www.wikipedia.org/>.
- [18] Ekstrand M., Riedl J. & Konstan J, *Collaborative Filtering Recommenders Systems Vol.4, No.2(2010), 81-173*

3D Facial Features in Neuro Fuzzy Model for Predictive Grading Of Childhood Autism

Reji R¹, Dr P SojanLal²

¹Research Scholar, School of Computer Sciences,

Mahatma Gandhi University, Kottayam, Kerala, India.
reji.r@saintgits.org

² Principal, Mar-Baselious Institute of Technology & Science,

Kothamangalam, Kerala, India.
sojanlal@gmail.com

Abstract: Autism Spectrum Disorder (ASD) is a clinically heterogeneous neurological developmental disorder. It is called a spectrum disorder because of its range of symptoms. Early diagnosis and proper intervention is required for the effective treatment of autism. Diagnosis is based on the quantitative and qualitative analysis made by the clinician. The expertise of the clinician is so important in the proper diagnosis and classification of autism. This paper proposes an Expert system that act as a support system to the clinician. Major clinical attributes of autism along with facial features are used as input to the expert system. The main highlight is the use of features from 3D facial imagery for autism classification. The expert system operates in two modes, diagnosis mode and grading mode. Naïve Bayes classifier is initially used for diagnosis mode where as overall system is implemented using a Neuro-Fuzzy approach. In the diagnosis mode 100% accuracy and in classification mode 98.8% accuracy is obtained.

Keyword: Autism, ASD, 3D Face, Neuro-Fuzzy System, Neural networks, Fuzzy logic, Expert system.

Introduction: Autism spectrum disorder (ASD) is a clinically heterogeneous condition with a wide range of factors. A satisfactory diagnosis measure for ASD is currently unavailable. Autism is a neurological handicap in children, which is usually diagnosed in early child hood. There is lack of definitive biomarkers for autism diagnosis. The diagnosis mostly depends on a range of factors. People with autism show different clinical features and symptoms. There is lot of scope for quantitative research on ASD in developing countries like India. Satisfactory and accurate data for research in autism is unavailable in India. The frequency of ASD diagnosis is increasing. Many Factors like increased awareness, improved detection mainly contribute to this. The publication in DSM-5 on May 2013 adds major revisions needed to remove the confusing labels associated with ASD. The earliest symptom is the absence of normal behavior. All children should be screened using a standardized Autism screening tool at 18 and 24 months of age [1].

Symptoms of ASD must be present in the earlier developmental period mostly by the second year of life (after 12 months). But least severe type of ASD may be diagnosed by 4 to 6 years or later. Intervention should begin as early as possible. In intervention consider the core distinctive features of autism and it should be specific and proof based. More over it should be well structured and appropriate to the developmental need of the child.

Even though there are inter individual difference in the clinical levels of children with ASD, they share some common characteristics like deficit in social interaction/communication and behavioral abnormalities.

Studies shows that children who deficit to recognize face in childhood shows severe autistic features at teenage. Researches shows that human recognize a person by their body if someone is coming from far away or the face is obscured. So for identifying a person, brain uses facial characteristics and also other physical cues.

Researchers at the University of Missouri have identified facial features measurements in children with autism and developed a screening tool for young children. The sample consists of children from 8 to 12 years of age. Judith Miles, Professor Emeritus of child health-genetics in the MU Thompson Center for Autism and Neuro Developmental Disorders point out that a portion of those children diagnosed with autism tend to look alike with similar facial characteristics [2].

In this research we are developing an expert system that use core clinical features with its attributes, facial characteristics and parental status as input.

Autism: Clinical features and Diagnosis

Autism detection can be done by using quantitative tests and qualitative analysis. In DSM IV ASD diagnosis is based on Language delays, Social Communication Problem and Repetitive behavior. Where as in DSM V ASD diagnosis is having two criteria domains namely Social interaction domain and Repetitive or restricted behavior domain. The Core Clinical features of autism can be brought under the following heads with attributes.

1. Behavioral problem
 - a) Poor eye contact
 - b) Lack of responsiveness to others
 - c) Difficulty in building social relationship
 - d) Repetitive acts
 - e) Self harm
 - f) Compulsive behavior
 - g) Hyper Activity
 - h) Poor joint attention
 - i) Solo play
 - j) Excessive fear
 - k) Poor emotional response

2. Language Disorder
 - a) Muteness
 - b) Echolalia
 - c) Sound making

3. Intellectual retardation
 - a) General intellectual retardation
 - b) Brain Seizures

4. Facial Features
 - a) Open Eyes
 - b) Wide Mouth
 - c) Large region between mouth and nose.
 - d) Expression less face
 - e) Open mouthed Appearance
 - f) Prominent Forehead

5. Parents Status
 - a) Not Autistic
 - b) Autistic

The earliest symptom is the absence of normal behavior. Normally when a parent or a healthcare provider notices any delay or abnormal behavior in the child at, or prior to the age of three they are prompted to consult a developmental pediatrician. The child is analyzed carefully and any abnormality is observed in the core functional areas, the developmental pediatrician recommends the child for assessment test using any of the standard autism testing tools. These tools are normally a checklist or questionnaire containing autism features. The clinician fills the data using his observation and a structured discussion with the parent of the child under scrutiny. After filling the details a final score is generated. Comparing the obtained score with the threshold value, the clinician initially classifies the child as either not autistic or autistic. The next step is to identify which Autistic class or grade the child belongs to. Based on the total score compared against a threshold the child is diagnosed as mild, moderate and severe. Consider the total score(S) adds up to 60 and the threshold is 30, the grade and remarks is as shown in table 1.

Score	Class/ Grade	Remarks
Score <30	Normal	Typical
Score 30 to 34	Mild	Requiring support
Score 34 to 38	Moderate	Requiring Substantial support
Score >38	Severe	Requiring very substantial support

Table 1: Score with Grade

The expertise and dedication of the clinician is an important factor while analyzing the grade or class of autism. Expert clinician can easily spot the grade of autism. Some clinician fully depends on the diagnosis tool and there are possibilities of wrong classification. More over the fuzziness in the Score may also lead to misclassification. Studies say that a proper initial diagnosis and follow up is required for autism. If we are using an expert system as a support system for clinicians the misclassification and problems in initial diagnosis of autism can be avoided up to an extent. In this research we are developing an expert system to assist clinicians in their diagnosis procedure.

Related work:

Silberberg et al.[3] focus on the prevalence of neuro-developmental disorder among children aged 2 to 9 years in the different areas of India. They also analyzed the risk factors associated with neuro-developmental disorders along with the development of screening and diagnosing methodology.

An investigation related to the epidemiology of ASD in India was reported by Mukerji et al.[4]

Myers et al.[5] suggests that the primary goal of treatment for ASD is to maximize the child's ultimate functional independence and quality of life by minimizing the core features of ASD.

Robins et al.[6] objective is to validate the modified checklist for Autism in toddlers.

Yasmin H. Nuggers[7] studied the prevalence, risk factors and diagnosis of ASD in developing countries. In his brief reviews controversies regarding the increase in estimate of prevalence, implications of changes in ASD definitions are also discussed.

Vijay Sagar KJ[8] focus on the study of developmental disorders in India. He concludes his article by saying that there is a need of proper diagnosis and screening tools for Autism in India.

Hammond et al.[9] proposes the use of dense face models in 3D Analysis of facial morphology. The model provide a detailed visualization of 3D face shape variation with capability to training the Physicians to recognize the core components of particular syndromes. Ten fold cross validation testing is done on the sample faces using different pattern recognition algorithm.

Vezzetti et al.[10] highlights 3D human face descriptions, land marks measures and geometrical features. Analysis of facial morphology is very important in the study of facial abnormalities.

Gupta et al.[11] worked on the assumption that different facial expressions can be considered as isometric deformation of facial surfaces .Even though deformation occurs, the intrinsic property of the surface remain the same.

Aldridge et al.[12] investigation focus mainly on the correlation between brain development and face. Brain develops in concert and coordination between the developing facial tissues. ASD is due to alteration in embryological brain, suggests that there are differences in the facial structures of ASD children and normally developing one. Finally the authors concludes that there are significant differences in the facial morphology of boys with a ASD compared normally developing one.

Weigelt et al.[13] reports the face identity recognition is deficit in ASD. The deficit is both process specific and domain specific. They suggest that Autism is a domain specific disorder.

Ruggeri et al.[14] objectives is to find the similarity and difference between the terms biomarker and endophenotype. There study includes the established biomarkers and endophenotype in autism research along with the discovery of new biomarkers.

Dataset: The background study and data collection for this work is done at Block Resource Centre Cherthala, Kerala, India. BRC is a Government agency working along with Sarva Shiksha Abhiyan. The dataset consists of 47 children, which includes both boys and girls. The ratio of boys and girls is 12: 1. The age is from 2 years to 12 years. While studying and analyzing the dataset we are making use of the expert opinion from Pediatric Neurologist, Developmental Pediatricians, Speech Therapist, Remedial Educators, Clinical Psychologist and Parents.

Objective: Our research focus on developing an expert system for the initial diagnosis and grading of childhood autism. This system can be used as a support system for the clinicians while diagnosing autism. The proposed system is having two modes of operation, Diagnosis mode and Grading mode as shown in figure 1. Initially in the diagnosis mode expert system predicts whether the child is non-autistic or autistic. Once the output of the diagnosis mode is autistic then the next phase is activated. In this phase a detailed analysis is done and the possible outcome is the class or grade of autism.

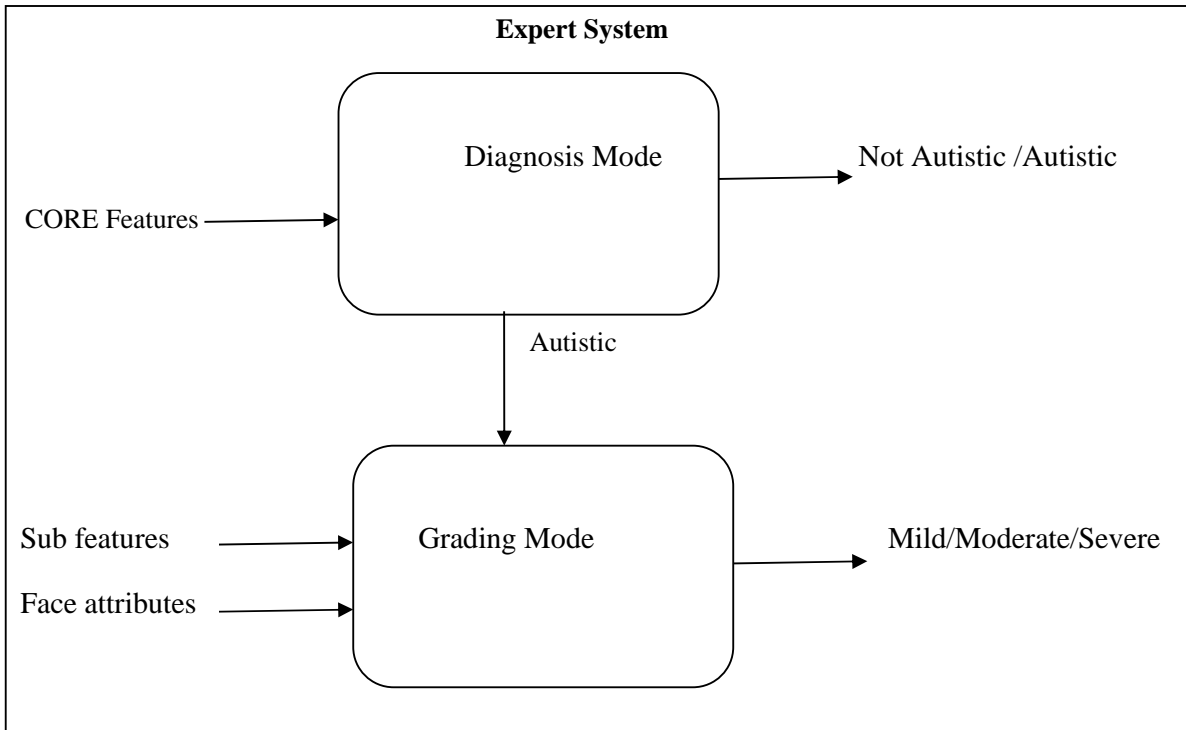


Figure 1: Flow chart of the expert system.

Scale	Output	Remarks
0	Normal	Non autistic
1	Mild	Requiring support
2	Moderate	Requiring Substantial support
3	Severe	Requiring very substantial support

Table 2: Grading

The core feature of autism is analyzed initially during the diagnosis mode. Core features includes Behavioral problem, Language disorder and General mental retardation. Based on the core features the diagnosis mode output is not autistic or autistic.

If the output is autistic then the second phase is activated, the sub features of each of the core features are analyzed. Facial features along with status of parents are along given importance. The most important characteristic of this expert system is the integration of facial features. The face image of the child under diagnosis is captured; it is modeled to 3D or captures the image using a 3D imaging system. In 3D imaging the geometric depth information is having more importance. Facial features include mouth, eyes and the region between mouth and nose. A portion of Autistic children's have wider mouth, open eyes and large region between the mouth and nose as shown in figure 4. Other common facial characteristics are expression less face, open mouthed appearance and prominent forehead region. Using 3D Geodesic distance as the measure identifies the variance of the features from normally developing kids. Our aim is to extract the exact geometrical information from the face under scrutiny and compare it with a template and used this information for training. By using these facial attributes our focus is to study the contribution of each feature to grading of autism.

Our analysis, point out the fact that Children below the age of 8 with other clinical features of Autism mostly lack the facial features mentioned above. But children from the age of 8 and 12 have shown the above mentioned facial features along with other clinical features of Autism. Our expert system is designed in such a way that the weightage of facial features is varied by considering the age of the child under diagnosis. If the age is below 8 the weightage of the features in percentage is as 75(core features) : 15(facial features) : 10 (Parents status). Whereas age range from 8 to 12 the weightage of the features in percentage is as 65(core features) : 25(facial features) : 10 (Parents status).Parent status is also considered, this feature include whether the parents are autistic or not and age of the parents during conception is also given weightage.

In the grading phase three sets of features namely attributes from core features, facial attributes and parent's status is considered. The weightage of the features varies depending on the age of the child. Initially we consider the two phases as two separate classification problem. In phase 1 the number of inputs are limited so a Naïve Bayes classifier is applied and it suites our problem and it gives the result autistic or non autistic as shown in figure 2. The input to the classifier is the core features such as Behavior problem, Language Disorder and General Mental retardation.

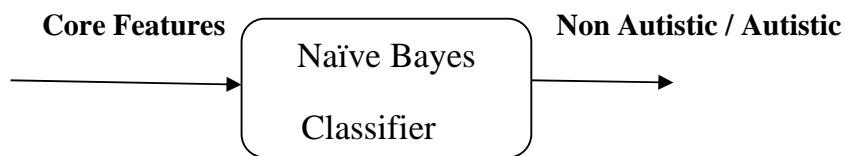


Figure 2: Diagnosis Mode

In the second phase more inputs belonging to different features are considered which include attributes from the core features, facial region and parental status. Naïve Bayes classifier is applied and result is analyzed but there exists some fuzziness after a certain threshold. We need to integrate the two phases and a neuro- fuzzy approach is applied. Soft computing approach like neural network and fuzzy logic can play a vital role in the design of such an expert system. Fuzzy logic is used to interpret expert knowledge directly using rules with linguistic base. In this system we are qualitatively collecting lot of information with structured discussion with parent and from clinician's observation.

Linguistic base can easily be framed into fuzzy rules. Neural network are good in recognizing patterns. So this hybrid approach yields better performance. The output of the grading phase is as shown in table 2.

Results:

To design the neuro-fuzzy system for diagnosis of autism we consider the attributes of core features, facial attributes and parental status. The hybrid architecture is as shown in figure 3.

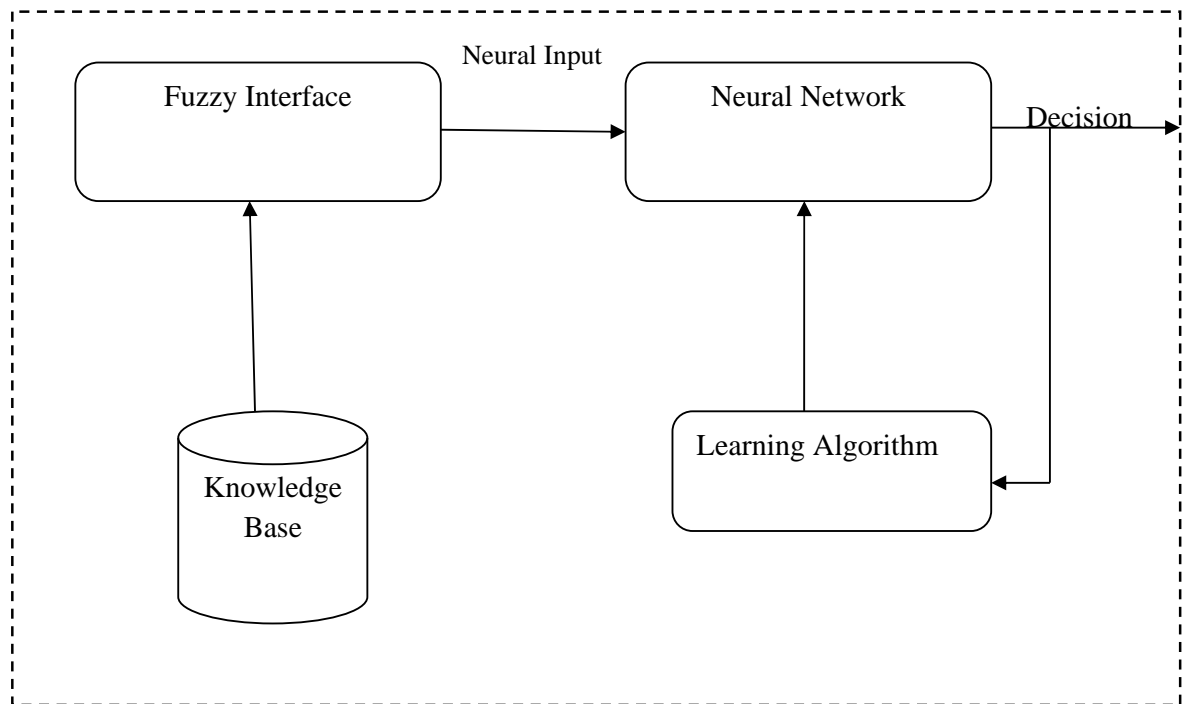


Figure3: Neuro-Fuzzy Expert System Architecture for predicting Autism

The knowledge base consists of twenty two fuzzy parameters. The neural network is trained to learn the parameters of the membership functions representing the linguistic terms in the rule. Sample fuzzy rules applied in the Knowledge base is as follows:

R1: If (Behavior Problem) && (Language Disorder) && (General Mental retardation) then belongs to class Autistic

R2: If (Behavior problem Attributes (1 || 2 ||n)) &&(Language Disorder Attributes (1 || 2 || n)) && (Mental retardation Attributes(1 || 2 ||n)) then belongs to class Autistic .

Different soft computing model have been tested like Naïve Bayes, SVM, K-Means, FCM and Neuro Fuzzy with the same input attributes using Weka tool .The performance is evaluated and the most outstanding results are shown in table4. The operational procedure of the neuro fuzzy system for autism classification is shown in figure 5

The expert system is tested and evaluated by the different stakeholders, the accuracy and evaluation survey summary is shown in figure 6 and 7.

Technique	Sample size	Inputs	Outputs	Accuracy rate
Naïve Bayes	47	12	2	100
Neuro- Fuzzy	47	22	4	98.8

Table4: Performance of Classifier.

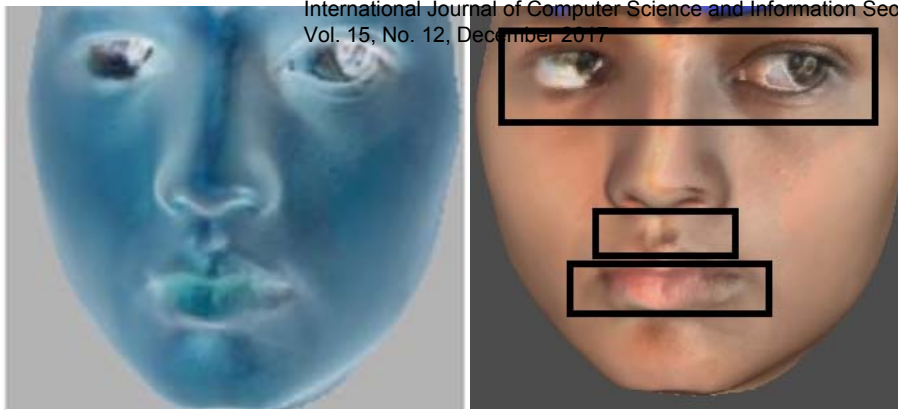


Figure 4: Facial Features

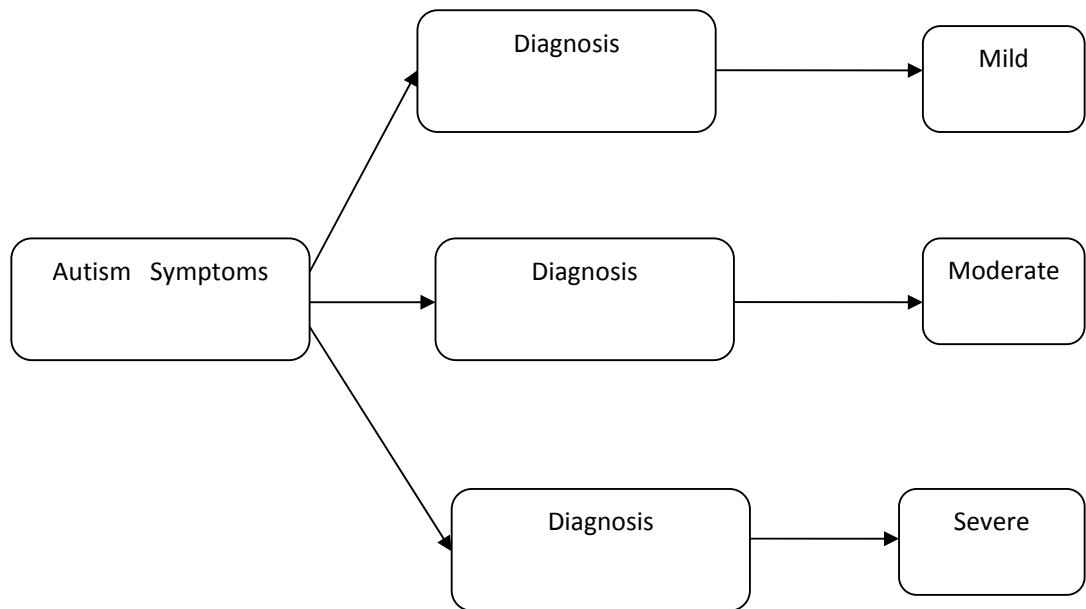


Figure 5: Operational Procedure of the Neuro-Fuzzy system for Autism classification

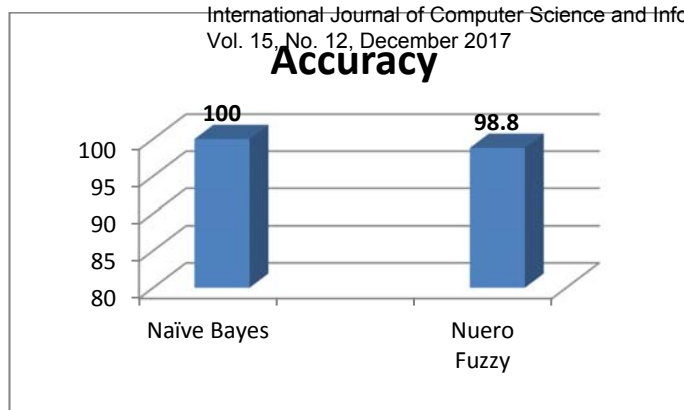


Figure: 6 Accuracy

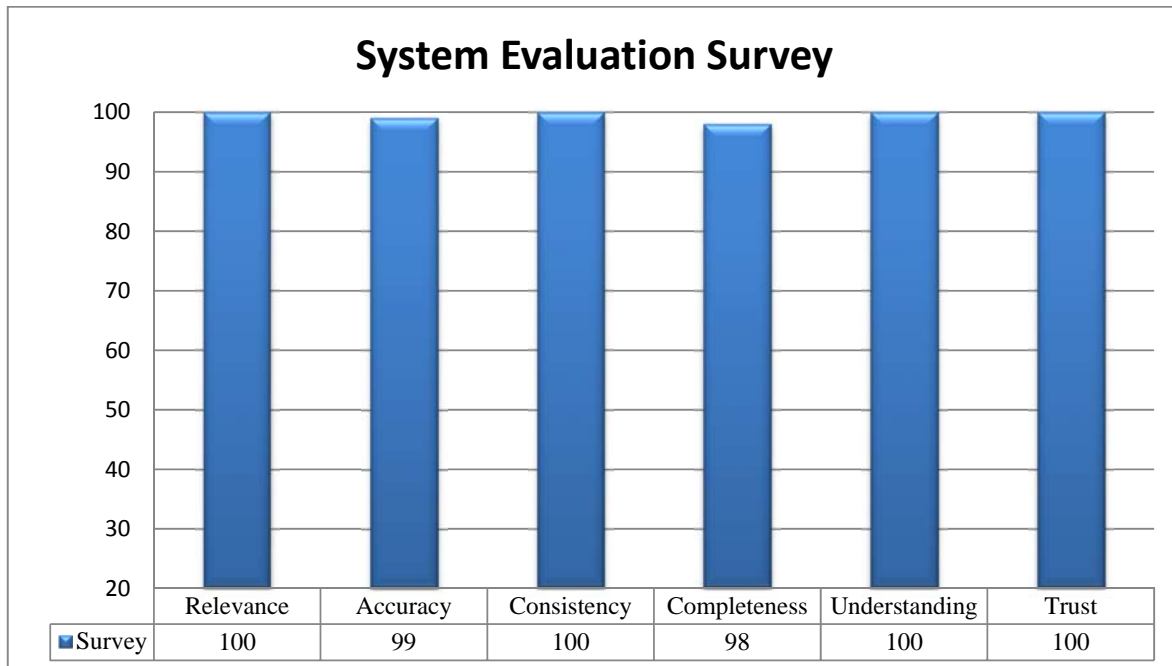


Figure 7: Expert system evaluation Survey

Conclusion: Studies related to the cause and symptoms of Autism spectrum disorder are going on around the world. Information Technology is finding lots of application in all fields. Due to the complexity and heterogeneous nature of this disorder, fewer works are reported which make use of IT in this area. Our expert system captures different inputs and produces an appropriate output. This system can be used by clinicians as a support system. The expert system is used and evaluated by 20 potential users and they all provide positive responses relating to input, output and quality of the system. Integrating 3D facial features as input to the system add a new dimension in Autism research.

- [1] Myers SM, Johnson CP. Management of children with autism spectrum disorders. *Pediatrics*.2007;120:1162-82.
- [2] Tayo Obafemi-Ajayi, Judith H. Miles, T. Nicole Takahashi, Wenchuan Qi, Kristina Aldridge, Minqi Zhang, Shi-Qing Xin, Ying He, Ye Duan. **Facial Structure Analysis Separates Autism Spectrum Disorders into Meaningful Clinical Subgroups**. *Journal of Autism and Developmental Disorders*, January 2015 DOI: 10.1007/s10803-014-2290-8
- [3]. Silberberg D, Arora N, Bhutani V, Durkin M, Gulati S, Nair M, *et al.* Neuro-Developmental Disorders in India-From Epidemiology to Public Policy. *Neurology*. 2014; 82:P7-P324.
- [4]. Mukerji S. A large scale, two phase study to estimate prevalence, and raise awareness, about autism spectrum disorders in India. Action for Autism, Jan 2009. New Delhi, India.
- [5]. Myers SM, Johnson CP. Management of children with autism spectrum disorders. *Pediatrics*.2007;120:1162-82.
- [6]. Robins DL, Casagrande K, Barton M, Chen CMA, Dumont-Mathieu T, Fein D. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHATR/F). *Pediatrics*.. 2014;133:37-45.
- [7]. Yasmin H. Neggers, Department of Human Nutrition, University of Alabama, Tuscaloosa AL, USA, Autism spectrum disorders: Increasing prevalence and changes in diagnostic criteria.
- [8]. Vijay Sagar KJ. Research on autism spectrum disorders in India. *AP J Psychol Med* 2011;12 (1): 69–72.
- [9]. Peter Hammond, Tim J. Hutton,1 Judith E. Allanson,Linda E. Campbell, Raoul C.M. Hennekam,Sean Holden, Michael A. Patton, Adam Shaw,I. Karen Temple, Matthew Trotter,Kieran C. Murphy, and Robin M. Winter, *American Journal of Medical Genetics* 126A:339–348 (2004)
- [10]. Vezzetti E.; Marcolin F. (2012). 3D human face description: landmarks measures and geometrical features. In: *IMAGE AND VISION COMPUTING*. - ISSN 0262-8856
- [11]. Gupta, S., Markey, M. K. , Bovic, A. C. (2010) “Anthropometric 3D Face Recognition”, *International Journal of Computer Vision*, Vol. 90, No. 3: 331-349.
- [12]. Aldridge, K., George, I. D., Cole, K. K., Austin, J. R., Takahashi, T. N., Duan, Y., & Miles, J. H. (2011). Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes. *Molecular Autism*, 2, 15. <http://doi.org/10.1186/2040-2392-2-15>
- [13]. Weigelt S, Koldewyn K, Kanwisher N (2013) Face Recognition Deficits in Autism Spectrum Disorders Are Both Domain Specific and Process Specific. *PLoS ONE* 8(9): e74541. doi:10.1371/journal.pone.0074541
- [14]. Barbara Ruggeri ,Ugis Sarkans , Gunter Schumann ,Antonio M. Persico, *Biomarkers in autism spectrum disorder: the old and the new*, Springer-Verlag Berlin Heidelberg 2013

Classification of Matrix Multiplication Methods Used to Encrypt-Decrypt Color Image

Saleh A. Khawatreh

Computer Engineering Dept. Faculty of Engineering

Al-Ahliyya Amman University

skhawatreh@ammanu.edu.jo

Abstract

The digital color images are the most important types of data is now circulating on the Internet, so the protection and security of the image transition has the top priorities of computer experts. Many researchers had developed different techniques to increase the security of image transmission and most of these techniques suffer from the slow of the encryption-decryption process. In this paper we will produce a classification of the most popular encryption-decryption techniques and suggest an efficient one, the suggestion will be based in many factors such as speedup, throughput, encryption-decryption error and the hacking factor.

Key words: Encryption, decryption, speedup, throughput, hacking.

1. Introduction

Encryption is defined as the conversion of plain message (matrix which represents digital color image) into a form called a cipher text that cannot be read without decrypting the encrypted text [15]. Decryption is the reverse process of encryption which is the process of converting the encrypted text into its original plain text, so that it can be read [15]. Color image encryption is to be done before transition the image

and it has to be done securely over the network so that no unauthorized user can able to decrypt the image. Image encryption, video encryption, chaos based encryption have applications in many fields including the Internet communication, transmission, medical imaging. Tele-medicine and military Communication, etc. The evolution of encryption is moving towards a future of endless possibilities. The image data have special properties such as bulk capability, high redundancy and high correlation among the pixels. Encryption techniques are very useful tools to protect secret information [3].

Encryption of data [16] has become an important way to protect data resources especially on the Internet, intranets and extranets. Encryption is the process of applying special mathematical algorithms and keys to transform digital data into cipher code before they are transmitted and decryption involves the application of mathematical algorithms and keys to get back the original data from cipher code. The main goal of security management is to provide authentication of users, integrity, accuracy and safety of data resources [16].

2. Related works

Guodong Ye [9] have presented an efficient image encryption algorithm

using double logistic maps, in which the digital matrix of the image is confused from row and column respectively. Confusion effect is carried out by the substitution stage and Chens system is employed to diffuse the gray value distribution. Haojiang Gao *et al.* [5] have presented an algorithm presented a Nonlinear Chaotic Algorithm (NCA) by using power and tangent functions instead of linear function. The encryption algorithm is a one-time-one-password system and is more secure than the DES algorithm. Jawahar Thakur *et al.* [17] presented a comparison between symmetric key algorithms such as DES, AES, and Blowfish. The parameters such as speed, block size, and key size are considered to evaluate the performance when different data loads are used. Blowfish has a better performance than other encryption algorithms and AES showed poor performance results compared to other algorithms due to more processing power.

Khaled Loukhaoukha *et al.* [9] introduced an image encryption algorithm based on Rubik's cube principle. The original image is scrambled using the principle of Rubik's cube and then XOR operator is applied to rows and columns of the scrambled image using two secret keys. Liu Hongjun *et al.* [18] designed a stream-cipher algorithm based on one-time keys and robust chaotic maps. The method uses a piecewise linear chaotic map as the generator of a pseudo-random key stream sequence.

M. Zeghid *et al.* [19] analyzed the AES algorithm, and added a key stream generator (A5/1, W7) to AES to ensure improved encryption performance

mainly for the images. The method overcomes the problem of textured zones existing in other known encryption algorithms. Maniccam *et al.* [20] presented a method for image and video encryption and the encryption methods are based on the SCAN methodology. The image encryption is performed by SCAN-based permutation of pixels and a substitution rule which together form an iterated product cipher. The pixel rearrangement is done by scanning keys and the pixel values are changed by substitution mechanism. Figure 1 shows the basic SCAN patterns used in [16]. Mohammad Ali el al. [21] introduced a block-based transformation algorithm based on the combination of image transformation and the Blowfish algorithm. The algorithm resulted in the best performance by the lowest correlation and the highest entropy. The characteristics of AES are its security and resistance against attacks and the major characteristic of RC4 algorithm is its speed [11]. A hybrid cipher by combining the characteristics of AES and RC4 is developed and 20% improvement in speed is achieved when compared to the original AES and a higher security compared to the original RC4 [13].

Rizvi *et al.* [12] analyzed the security issues of two symmetric cryptographic algorithms Blowfish and CAST algorithm and then compared the efficiency for encrypting text, image, and audio with the AES algorithm across different widely used Operating Systems. For text data, all algorithms run faster on Windows XP but Blowfish is the most efficient and CAST run slower than AES. Blowfish encrypts images most efficiently on all the three

platforms. For audio files, CAST performs better than Blowfish and AES on Windows XP but on Windows Vista and Windows 7, there is no significant difference in the performance of CAST and AES; however, Blowfish encrypts audio files at less speed.

Sanfu Wang *et al.* [21] presented an image scrambling method based on folding transform to folding matrix which is orthogonal and enables to fold images either up-down or left-right. When an image is folded this way repeatedly, it becomes scrambled. The scrambling algorithm has an effective hiding ability with small computation burdens as well as wide adaptability to images with different scales.

Sathishkumar G.A *et al.* [14] presented a pixel shuffling, base 64 encoding based algorithm which is a combination of block permutation, pixel permutation, and value transformation. The crypto system uses a simple chaotic map for key generation and a logistic map was used to generate a pseudo random bit sequence. The total key length is 512 bits for each round and the key space is approximately 2512 for ten rounds. Shao Liping *et al.* [4] proposed a scrambling algorithm based on random shuffling strategy which could scramble non equilateral images and has a low cost to build coordinate shifting path. The algorithm is based on permuting pixel coordinates and it could be used to scramble or recover image in real time. T. Sivakumar, and R. Venkatesan [4] proposed a novel image encryption approach using matrix reordering this approach was tested and some comparisons with other techniques were done.

Ziad A. Alqadi and others in [1] and [2] have presented a technique using direct and inverse conversions to convert a color image to gray image and vice versa, this technique can be useful to be used in color image encryption decryption.

3. Proposed methods

3-1 First method (proposed 1): Using each of the components of the color image

This method for encryption can be implemented in the following steps:

1. Get the original color image.
2. Extract the red, green, and blue matrices from the original color image (each of them is 2 dimensional matrix),
3. Reshape each matrix in step 2 to square matrix.
4. Generate one random square matrix for each component to be used as a private key.
5. Encrypt each component by applying matrix multiplication of the matrix component and its private key.
6. Reshape each encrypted matrix to its original size.
7. Form the encrypted color image.

The decryption phase can be implemented applying the following steps:

1. Get the decrypted color image.
2. Extract the red, green, and blue matrices from the original color

- image (each of them is 2 dimensional matrix),
3. Reshape each matrix in step 2 to square matrix.
 4. Use each private key
 5. Decrypt each component by applying matrix multiplication of the matrix component and the inverse it's private key.
 6. Reshape each decrypted matrix to its original size.
 7. Form the decrypted color image.

The following matlab code was written to implement this method

```
clear all
close all
a=imread('C:\Users\User\Desktop\flower-color-combinations.jpg');
subplot(2,2,1)
imshow(a), title 'Original image'
subplot(2,2,2)
imhist(a(:,1)), title 'Red component histogram'
subplot(2,2,3)
imhist(a(:,2)), title 'Green component histogram'
subplot(2,2,4)
imhist(a(:,3)), title 'Blue component histogram'
tic
b1=a(:,1);
b2=a(:,2);
b3=a(:,3);
b1=reshape(b1,200*300,1);
b2=reshape(b2,200*300,1);
b3=reshape(b3,200*300,1);
for i=60001:60025
    b1(i,1)=0;
    b2(i,1)=0;
    b3(i,1)=0;
end
c1=reshape(b1,245,245);
```

```
c2=reshape(b2,245,245);
c3=reshape(b3,245,245);
k1=rand(245,245);
k2=rand(245,245);
k3=rand(245,245);
c1=double(c1);
c2=double(c2);
c3=double(c3);
e1=c1*k1;
e2=c2*k2;
e3=c3*k3;
toc
tic
d1=e1*inv(k1);
d2=e2*inv(k2);
d3=e3*inv(k3);
d11=reshape(d1,245*245,1);
d12=reshape(d2,245*245,1);
d13=reshape(d3,245*245,1);
for i=1:60000
    d21(i,1)=d11(i,1);
    d22(i,1)=d12(i,1);
    d23(i,1)=d13(i,1);
end
d31=uint8(d21);
d32=uint8(d22);
d33=uint8(d23);
d41=reshape(d31,200,300);
d42=reshape(d32,200,300);
d43=reshape(d33,200,300);
d4(:,1)=d41;
d4(:,2)=d42;
d4(:,3)=d43;
toc
figure
subplot(2,2,1)
imshow(d4), title 'Decrypted image'
subplot(2,2,2)
imhist(d4(:,1)), title 'Decrypted red component histogram'
subplot(2,2,3)
imhist(d4(:,2)), title 'Decrypted green component histogram'
subplot(2,2,4)
imhist(d4(:,3)), title 'Decrypted blue component histogram'
```


3-2 Second method (proposed 1): Converting color image to 2 dimensional matrix

The encryption phase here is consisted of the following steps:

1. Get the original digital color image as a 3 dimensional matrix(m).
2. Reshape m into 1 column matrix(r).
3. Get the size of r (s).
4. If s is a square number proceed to step 6.
5. Find the nearest square number to s and adjust s to this number, adjust r by padding zeros.
6. Reshape r to square matrix (r1).
7. Generate a double random square matrix with size equal r1 size, this matrix will be used as a private key for encryption-decryption (k).
8. Save k to be used in the decryption phase.
9. Get the encrypted image (e) by applying matrix multiplication of r1 and k.
10. Reshape e into 1 column matrix (e1).
11. Omit the padded zeros from e1.
12. Reshape e1 into 3 dimensional matrix to get the encrypted color image.

The decryption phase can be implemented applying the following steps:

1. Get the encrypted digital color image as a 3 dimensional matrix(en1).
2. Reshape en into 1 column matrix(en2).
3. Get the size of en2 (s).
4. If s is a square number proceed to step 6.
5. Find the nearest square number to s and adjust s to this number, adjust en2 by padding zeros.
6. Reshape en2 to square matrix (en3).
7. Use the private key k.
8. Get the decrypted image (di) by applying matrix multiplication of r1 and the inverse of k.
9. Reshape di into 1 column matrix (di1).
10. Omit the padded zeros from di1.
11. Reshape di1 into 3 dimensional matrix to get the decrypted original color image.

The following matlab code was written to implement this method

```
clear all
close all
a=imread('C:\Users\User\Desktop\flower-color-combinations.jpg');
subplot(2,2,1)
imshow(a), title 'Original image'
subplot(2,2,2)
imhist(a(:, :, 1)), title 'Red component histogram'
subplot(2,2,3)
```

```
imhist(a(:,:,2)), title 'Green component  
histogram'  
subplot(2,2,4)  
imhist(a(:,:,3)), title 'Blue component  
histogram'  
tic  
b=reshape(a,200*300*3,1);  
for i=180001:180625  
    b(i,1)=0;  
end  
c=reshape(b,425,425);  
k=rand(425,425);  
c=double(c);  
e=c*k;  
toc  
tic  
d=e*inv(k);  
d1=reshape(d,425*425,1);  
for i=1:180000  
    d2(i,1)=d1(i,1);  
end  
d3=uint8(d2);  
d4=reshape(d3,200,300,3);  
toc  
figure  
subplot(2,2,1)  
imshow(d4), title 'Decrypted image'  
subplot(2,2,2)  
imhist(d4(:,:,1)), title 'Decrypted red  
component histogram'  
subplot(2,2,3)  
imhist(d4(:,:,2)), title 'Decrypted green  
component histogram'  
subplot(2,2,4)  
imhist(d4(:,:,3)), title 'Decrypted blue  
component histogram'
```

3-3 Third method (proposed 1):Converting color image to Gray image

This method can be implemented as first method but the color image is to converted to gray image using direct conversion proposed by the author in [1], then the gray image can be encrypted as in method 1, after that the encrypted

gray image can be decrypted and converted to color image using the inverse conversion mentioned in [1].

4. Experimental results

The proposed methods were implemented several times using different color images with different sizes and the results always give a correlation coefficient equal 1 between the original image and the decrypted one, which means that the methods are 100% correct and do not lead to any damage of information, figure 1 and 2 show the original image and the decrypted one with the histogram of each component of the color image.

The proposed method is also very secure and it is implausible to hack the image because the private key has the following features:

- Private key is a 2 dimensional matrix with a huge size.
- Each element in the private key is a random double number which make it impossible to guess.

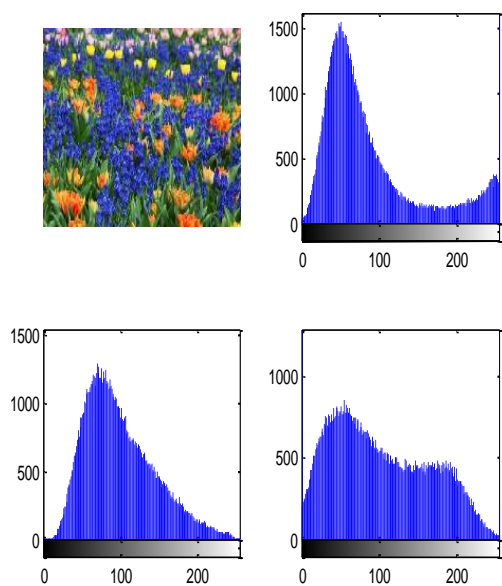


Figure 1: sample of the original color image

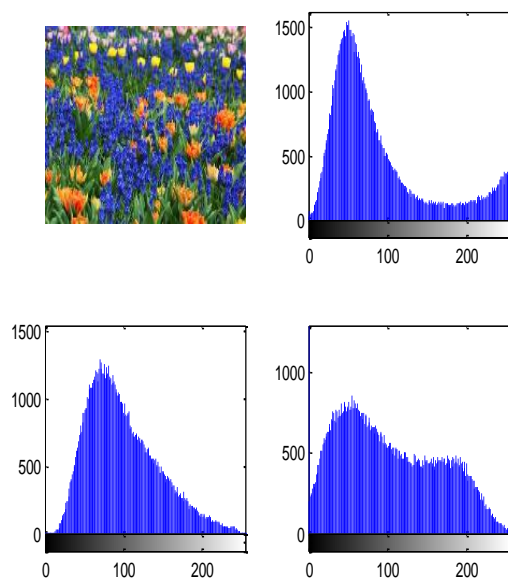


Figure 2: Decrypted color image.

The encryption and decryption times were calculated and compared with other methods mentioned in the related works, these results are listed in table 1.

Table 1: Comparisons results

Method	Direct conversion time (s)	Inverse conversion time (s)	Encryption time (s)	Decryption time (s)	Total time (s)	Speed up	Throughput (MB/s)
Proposed 1	0	0	0.006207	0.067798	0.0740	1	21.2549
Proposed 2	0	0	0.027985	0.156311	0.1843	2.4905	8.5343
Ref[1] HSI	0.078	0.032	0.02032	0.02541	0.1557	2.1041	10.1019
REF [1] R'G'B'	0.015	0.015	0.02032	0.02541	0.0757	1.0230	20.7776
Ref.[4]	0	0	0.23	0.23	0.46	6.2162	3.4193
Ref.[5]	0	0	0.5	0.5	1.0	13.5135	1.5729
Ref.[6]	0	0	0.12	0.12	0.24	3.2432	6.5536
Ref.[7], (A-I)	0	0	0.56	0.56	1.12	15.1351	1.4043
Ref.[7], (A-II)	0	0	1.01	1.01	2.02	27.2973	0.7786
Ref.[8]	0	0	0.4	0.4	0.8	10.8108	1.9661

The speedup was calculated by dividing the total time of the method by the total time of proposed 1 (which was taken as a reference because it has the best efficiency).

The throughput was calculated by dividing the color image size by the total time.

For clarity we can represent the data in table 1 by figures 3 and 4.

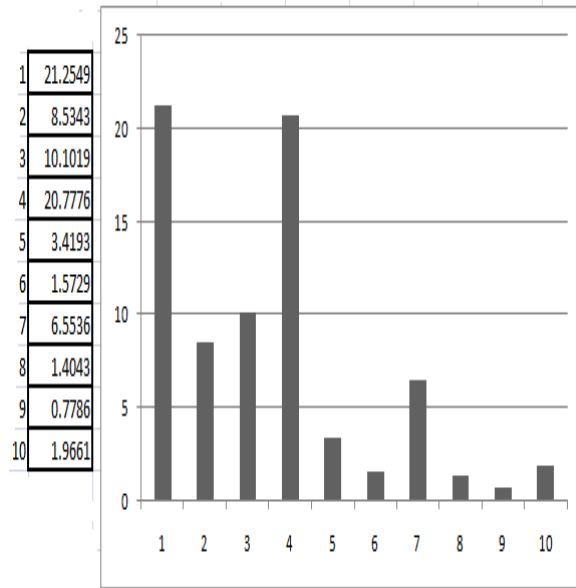


Figure 3: Methods throughput

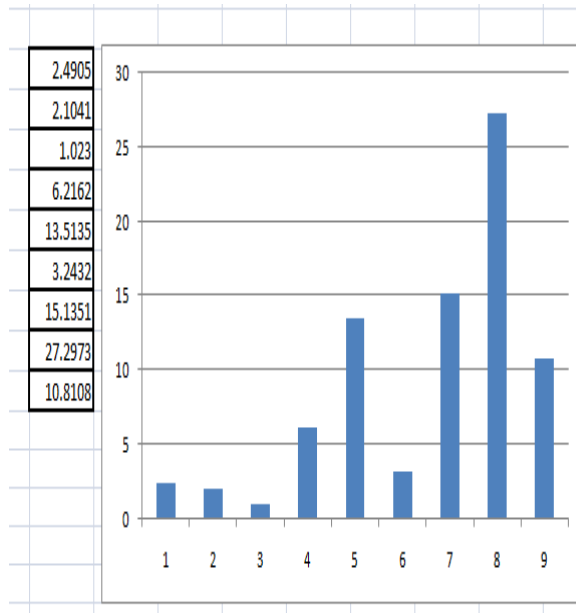


Figure 4: Speedup of the method reference to proposal 1

From the above results we can see that the proposal method 1 has the best efficiency.

Conclusions

A methods of encryption-decryption of color image were proposed and a survey analysis was done and it was shown that proposed 1 method has the best performance because it characterized with following features:

- Best speed in encryption phase.
- Best speed in decryption phase.
- Best throughput.
- No any damage of information.
- Impossible to hack.

References

[1]: Majed O. Al-Dwairi, Ziad A. Alqadi, Amjad A. AbuJazar and Rushdi Abu Zneit, Optimized True-Color Image Processing, World Applied Sciences Journal 8 (10): 1175-1182, 2010 ISSN 1818-4952.

[2]: Waheeb, A. and Ziad AlQadi, 2009. Gray image reconstruction. Eur. J. Sci. Res., 27: 167-173.

[3]: Rojo, M.G., G.B. García, C.P. Mateos, J.G. García and M.C. Vicente, 2006. Critical comparison of 31 commercially available digital slide systems in pathology. Int. J. Surg. Pathol., 14: 285-305.

[4]: T. Sivakumar, and R.Venkatesan , A Novel Image Encryption Approach using Matrix Reordering, WSEAS TRANSACTIONS on COMPUTERS,

Issue 11, Volume 12, November 2013,pp
407-418.

[5]: Haojiang Gao, Yisheng Zhang, Shuyun Liang and Dequn Li, "A New Chaotic Algorithm for Image Encryption", *Elsevier Science Direct*, vol. 29, no. 2, 2006, pp.393-399.

[6]: Khaled Loukhaoukha, Jean-Yves Chouinard, and Abdellah Berdai, "A Secure Image Encryption Algorithm Based on Rubik's Cube Principle", *Journal of Electrical and Computer Engineering*, 2011, pp. pp.1-13.

[7]: Xiaomin Wang, and Jiashu Zhang, "An Image Scrambling Encryption using Chaos- controlled Poker Shuffle Operation", *IEEE International Symposium on Biometrics and Security Technologies*, Islamabad, 23-24 April 2008, pp.1-6.

[8]: G. Chen, Y. Mao, and C. K. Chui, "A Symmetric Image Encryption Scheme based on 3D Chaotic Cat Maps", *Chaos, Solitons and Fractals*, Vol. 21, No. 3, 2004, pp.749–761.

[9]: Guodong Ye, "An Efficient Image Encryption Scheme based on Logistic maps", *International Journal of Pure and Applied Mathematics*, Vol.55, No.1,2009, pp. 37-47.

[10] Han Shuihua and Yang Shuangyuan, "An Asymmetric Image Encryption Based on Matrix Transformation", *ECTI Transactions on Computer and Information Technology*, Vol.1, No.2, 2005, pp. pp.126-133.

[11] Prabhudesai Keval Ketan and Vijayarajan V, "An Amalgam Approach using AES and RC4 Algorithms for

Encryption and Decryption", *International Journal of Computer Applications*, Vol.54, No.12, 2012, pp.29-36.

[12]: S.A.M Rizvi, Syed Zeeshan Hussain and Neeta Wadhwa, "A Comparative Study of Two Symmetric Encryption Algorithms Across Different Platforms", *International Conference on Security and Management (SAM'11)*, World Academy of Science, USA, 2011.

[13] Sanfu Wang, Yuying Zheng and Zhongshe Gao, "A New Image Scrambling Method through Folding Transform", *IEEE International Conference on Computer Application and System Modeling*, Taiyuan, 22-24 Oct. 2010, pp. v2-395-399.

[14] G.A. Sathishkumar and K. Bhoopathy Bagan, "A Novel Image Encryption Algorithm Using Pixel Shuffling and BASE 64 Encoding Based Chaotic Block Cipher", *WSEAS Transactions on Computers*, Vol.10, No. 6, 2011, pp. 169-178.

[15] John Justin M, Manimurugan S, "A Survey on Various Encryption Techniques", *International Journal of Soft Computing and Engineering (IJSCE)* ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[16] Ephim M, Judy Ann Joy and N. A. Vasanthi, "Survey of Chaos based Image Encryption and Decryption Techniques", *Amrita International Conference of Women in Computing (AICWIC'13)* Proceedings published by International Journal of Computer Applications (IJCA).

[17]: Jawahar Thakur, and Nagesh Kumar, "DES, AES, and Blowfish:

Symmetric Key Cryptography Algorithms Simulation Based Performance Analysis”, *International Journal of Emerging Technology and Advanced Engineering*, Vol.1, No.2, 2011, pp.6-12.

[18]: Liu Hongjun and Wang Xingyuan, “Color image encryption based on one-time keys and robust chaotic maps”, *Journal of Computers and Mathematics with Applications (Elsevier)*, Vol.59, 2010, pp. 3320-3327.

[19]: M. Zeghid, M. Machhout, L. Khriji, A. Baganne, and R. Tourki, “A Modified AES Based Algorithm for Image Encryption”, *World Academy of Science, Engineering and Technology*, Vol.3, 2007, pp.526-531.

[20]: S.S. Maniccam, and N.G. Bourbakis “Image and Video Encryption using SCAN Patterns”, *The Journal of the Pattern Recognition Society*, Vol.37, 2004, pp.725-737.

[21]: Mohammad Ali Bani Younes and Aman Jantan, “Image Encryption using Block-Based Transformation Algorithm”, *IAENG International Journal of Computer Science*, Vol.35, No.1, 2008, pp.3-11.

Document Processing and Arabic Optical Character Recognition: A User Perspective

Yasser M. Alginahi, Senior Member, *IEEE*

Abstract— The technological era we live in has influenced our daily activities at home, work and everywhere. The software technology has prospered, and new technology is being introduced to end-users every day. Document image analysis software has been influenced with more applications being developed. Hence, this inspired us to perform a general study on application software with emphasis on document analysis software i.e., Optical Character Recognition (OCR). This study starts with a brief review on recent works on document processing and Arabic OCR. Following this, a questionnaire survey is conducted to investigate the capability and familiarity of individuals with the top four Arabic OCR software, in order to provide recommendations and future research directions in this area. The results show that OCR is an essential technology which should be available with operating system tools. From the survey results, many of the respondents are not familiar with the top four commercial OCR, even those who are in the Information Technology (IT) sector. The study concludes that the available commercial OCR software are becoming more efficient; however, the accuracy rate reported needs to be further evaluated to provide a more accurate performance/recognition rate and, further investigation is needed to analyze the reported commercial accuracy rates. Finally, the study concludes with recommendations and future research directions.

Index Terms— Character recognition, Document Analysis, Editing Software, Optical Character Recognition Software, Scanner.

I. INTRODUCTION

DOCUMENT processing is an area of research that includes pattern recognition, artificial intelligence, data mining, information retrieval, image processing, and computer vision. Document processing, i.e. OCR, has been used in coordinating and conducting business transactions, mail sorting, check processing, passport processing, online publishing, digital libraries ... etc. The advancements in technology has revolutionized document processing, with the increase in speed, emergence of new storage medium, and increase storage capacity. Document processing is the capture of information from a paper medium into a digital medium, in other words, digitization. This may involve processing documents containing text only or documents containing mixed content (text and images). The process starts by separating/segmenting the documents into text and images which are then processed separately. Nowadays, with the available advanced technology, documents can be processed automatically with high precision results. Text processing involves the conversion of paper or image documents into digital formats, i.e., OCR, which can be further processed. OCR is a common method of digitizing

printed texts, to enable searching, editing, storing, transmitting or further processing of documents [1]. Similarly, image processing involves editing, storing, transmitting and further processing of images.

The term digitization has emerged since the last two decades. It is the conversion of paper documents into electronic formats, and nowadays the terms e-government, e-library, e-services, e-learning, e-commerce ... etc. have emerged; meaning digitization is being applied in many sectors/areas. Hence, the reduction in the amount of papers used is seen in many areas. Many libraries are reducing the paper books on their shelves and converting to digital libraries. Therefore, with the vast amount of information to be digitized many governments, firms and libraries around the world have started projects to digitize their paper contents, especially ancient manuscripts, thesis, books, and old documents. Therefore, digital libraries are now available online, which makes it easier to find information on a click of a button [2].

Computers are becoming faster and more reliable in document processing; in addition, database technologies for archived information maintenance have existed for long time and been used to store vast amount of information. Thus, OCR intelligence can be applied to digital image documents, which means the reduction in cost and time, since searching for information on a computer is much faster than finding it in a pile of dusty documents, which may take from several minutes to several hours. Therefore, the purpose of digitization is to produce digital documents which can be edited, stored, searched, transmitted online and used in other applications such as machine translation, text mining, and text-to-speech conversion.

This study explores the user experience with application software, document processing software and the four top available Arabic OCR software. The purpose is to get the user's perspective on document processing software. The main objectives of this study are:

- To explore the level of competence of individuals with basic application software skills in creating and sharing documents.
- To investigate how familiar and capable individuals are with OCR software.
- To investigate the accuracy of the recognition rate of the top four commercial OCR software from the user perspective.
- To study the current status of research in the area of Arabic OCR.

This paper is organized as follows. After this introduction, Section 2 provides the literature survey. Section 3 explains the research methodology. Section 4 presents the results and analysis. Section 5 provides the limitations and discussion, and

finally the conclusion and future work are stated in section 6.

II. LITERATURE REVIEW

The efficiency of document analysis has progressed and advanced with technology. Applications which use OCR have increased with the emergence of more efficient OCR software. The early versions of OCR software were very limited and worked for specific applications on specific fonts at a time. Early OCR may be traced to technologies involving telegraphy and creating reading devices for the blind [3], such as the machine developed by Emmanuel Goldberg in 1912 which converts characters into standard telegraph code and the Optophone device developed by Fournier D'Albe in 1914. Developments in OCR continued from the early 20th century until the era of the first digital computer in the early 1940s. Then, in the 1950s OCR became part of the business world [4]. Nowadays, advanced systems can produce a high degree of recognition accuracy for most fonts and different file formats, providing results which are similar to the format and layout of the input documents including columns, images, fonts, styles ... etc. Currently, OCR is available online as a service in a cloud-computing environment i.e., APIs, which can be used from any device connected to the Internet. These OCR API provide a simple way of parsing images and even multi-page documents to provide the results as a text file. In addition, various commercial and open source OCR systems are available for most of the common languages [5 – 6].

The work on document analysis is very active. Recently, the shift in research has moved from segmenting simple documents with text and images to segmenting text from scenes, billiard boards, movies ... etc. Document analysis goes through many steps and starts with processing the document digital image and ends with characters being recognized. The document analysis starts with a digital document image, then preprocessing is applied depending on the quality of the image. Preprocessing may include filtering noise, de-skew image, converting to gray or binary image ... etc. Next, page segmentation techniques are applied to identify different regions on the page, this is shown in Figure 1. After this, the text and graphics/images regions are both processed separately. Processing of the text portion of a document image includes: script recognition, font recognition, line/paragraph segmentation and word segmentation, Figure 2. Further processing of word segmentation includes, Holistic Word Recognition, Integrated Segmentation and Recognition, and Character Segmentation (OCR), Figure 3.

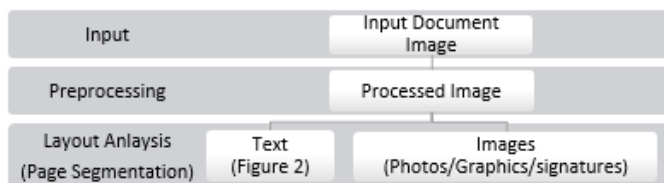


Figure 1: Document Image Analysis (Page Segmentation)

A general OCR system, Figure 3, goes through the following stages: preprocessing, character segmentation, feature extraction, classification and recognition. After a digital image

is fed into the system, preprocessing techniques may be used to remove noise, de-skew image, sharpen images ... etc.

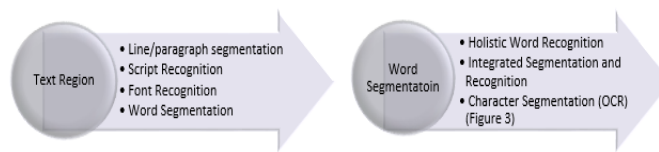


Figure 2: Text Region Processing

Next, the characters are isolated or segmented, and normalized. In addition, other techniques may be used to help prepare the characters for the feature extraction stage. The next stage is to classify and recognize the characters. Finally, post processing (contextual, grammatical information or data dictionary authentication), may be used to aid in achieving higher recognition rates. All these stages work in a pipeline fashion. The previous stage feeds into the next stage; therefore, the success of each stage guarantees an efficient OCR system.

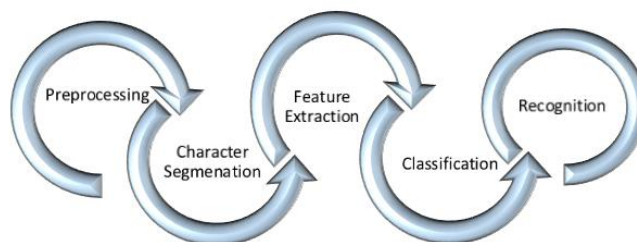


Figure 3: General OCR System

The work on OCR is still active in all stages presented in Figure 3. New techniques for OCR preprocessing, feature extraction, classification, and recognition are being published in recent literature. The work in [7] presents a comprehensive survey on character segmentation and challenges for segmentation of Arabic script based languages, i.e, Arabic, Urdu, Persian, Pashto, Sindi and Malay (Jawi). However, this research concentrated more on Urdu and concluded that the cursive nature of the Arabic script is the main challenge in character segmentation especially in Nasta'liq writing style compared to Naskh. Some other works related to document analysis include analysis of documents with non-uniform background [8], page segmentation [9], Arabic script recognition [10] [11], and segmentation of Arabic characters [12]. In addition, recent surveys in the areas of printed and handwritten Arabic and Urdu OCR research are available in the literature [13 – 17].

The research work on OCR for cursive scripts, such as Arabic, are not as mature as is the case with Latin scripts. Therefore, document image databases are needed for training, testing, validating and eliminating of errors. The work in [18] presents a multilingual image database created from various texts containing multiple fonts collected from various sources and 84 language scripts including Arabic script. The Arabic script is adopted by languages such as Kashmiri, Kurdish, Pashto, Persian, Punjabi (Shahmukhi), Sindhi, Urdu, and Uyghur. The images in the database were converted into word,

single line, multiline and paragraph images. Therefore, this database provides a platform for a step towards the establishment of standardized image database for the OCRs of the world language scripts. This database can be expanded to include more images and scripts, which researchers in the area of document analysis can benefit from [18]. Other works which discussed databases include the following scripts: Sindi [19], handwritten Persian [20], Pashto [21], handwritten Urdu [22], and Arabic [23]

Recent OCR works on printed Arabic scripts include the works done on the following Arabic scripts Urdu [24], Pashto [25], Persian [26], Punjabi [27], Uyghur [28], and Sindhi [29]; and for handwritten Arabic scripts include Urdu [30], Pashto [31], Persian [32], Punjabi [33], Uyghur [34], and Sindhi [35].

The research in the area of printed Arabic OCR resulted in a number of high quality commercial recognition systems. However, the general problems of such systems have not been solved yet. The main challenges, which are not solved in many OCR systems, still linger: multi-language/script recognition, multi-column reading order, very noisy documents, layout retention, and handwriting script [36]. The authors in [36] developed a multi-stage approach to document analysis involving preprocessing, content segmentation, recognition, and correction.

Currently, commercial OCR software for printed Latin scripts have reported 100% accuracy rates; however, this rate is not possible without the use of contextual, grammatical information or data dictionary authentication even where clear imaging is available. For example, using a smaller dictionary can help achieve a higher recognition rate for reading the amount line of a cheque. As a result, the recognizing of words from a dictionary is easier than trying to parse individual characters from script [37]. One study based on recognition of newspaper pages from the 19th and early 20th century concluded that character-by-character OCR accuracy for commercial OCR software varied from 81% to 99% [38]. On the other hand, the accuracy rates for hand-printed text is 80% to 90% on neat, clean hand-printed characters, and hence this accuracy rate still translates to dozens of errors per page, making the technology useful only in very limited applications. Furthermore, the accuracy rates for cursive text is still low especially in scripts other than Latin where more research is needed to achieve higher rates. Recognition of hand-printing, cursive handwriting, and printed text in other scripts are still the subject of active research [37 – 39].

OCR is used to be part of scanner software; then, it was available to be used on image documents without the use of a scanner. Now, OCR software is available as mobile apps where documents are being processed by taking an image using the mobile camera. Also, some OCR companies are providing OCR as an online service, where the documents are being processed on the cloud, i.e., OCR API [6]. Therefore, OCR applications are real-time and are used in government offices for scanning passport applications, license plate identification, ... etc. Nowadays, several OCR companies provide online service to process documents on cloud [6]. In this study, the top four Arabic commercial OCR software are studied in order to

provide recommendations and future directions in the area of OCR; these software are, Adobe Acrobat, OmniPage Standard, ABBYY FineReader, and Readiris. Table 1 shows the OCR performance rates and number of languages supported for these software. The data in Table 1 is obtained from www.toptenreviews.com [40].

Table 1: Reported OCR performance rates and number of languages supported

	Usability	Text Accuracy	Languages Recognized
Adobe Acrobat	91%	100%	190
OmniPage Standard	83%	99.80%	120
ABBYY FineReader	66%	99.84%	190
Readiris	75%	99.83%	130

*source: <http://www.toptenreviews.com/business/software/best-ocr-software/> [40]

Most of these software report perfect or close to perfect accuracy rates; however, no specific accuracy rates are provided for each of the languages supported by the software. Therefore, further studies are needed to evaluate the recognition rates of each language. As it can be seen, the highest number of languages supported is 190 by three software namely: Adobe Acrobat, and ABBYY FineReader.

A simple search for the term “OCR” on www.download.com provides 503 hits as of 24th of April 2017, between open source and commercial software [41]. Most of these software are open sources and they are usually research projects implemented by individuals or research groups. In general, the performance of open source OCR software is usually very low, have few editing capabilities, and is far from reaching the performance levels of commercial OCR software [41].

III. RESEARCH METHODOLOGY

A semi-structured questionnaire survey was designed. The survey aims to investigate the objectives outlined in the previous section. The questionnaire was first prepared and then disseminated to five subjects. After that, the questionnaire was modified. The survey questions are given in Table 2. Then three semi-structured interviews with faculty members from the IT sector were conducted to discuss the survey questions. The purpose of this step is to get feedback and comments on the clarity of the survey in order to consolidate some of the findings that were observed from the survey. Then, after an in-depth review of the subject matter, the semi-structured questionnaire was designed using www.monkeysurvey.com as a tool for data collection.

The survey was disseminated to the author’s contacts using email, WhatsApp, and Researchgate contacts. These individuals are mainly working in higher-level academic institutions all over the world. Approximately, 800 emails were sent, and it is expected that the response rate will be between 10 – 20%. This study explores the user experience with application software, document processing software and the four top Arabic OCR software in order to investigate the user’s perspective on document analysis software.

Table 2: The survey questions with no. of responses to each question

No.	Question Statement	Type of Question	Answered	Skipped	Percentage Answered
1	In what country do you live?*	Dropdown	124	2	98.4%
2	What is your gender?	Multiple Choice	124	2	98.4%
3	What is your age?	Dropdown	126	0	100.0%
4	What is your area of expertise/study?*	Dropdown	125	1	99.2%
5	What is the highest level of education you have completed?	Dropdown	125	1	99.2%
6	How important to your learning is it to have access to technology?	Multiple Choice	125	1	99.2%
7	How capable are you with each of the following application software to create and share documents?	Matrix/ rating scale	125	1	99.2%
8	How often do you use the following application software?	Matrix/ rating scale	125	1	99.2%
9	Which of the following problems you have encountered with the application software in Question 8? *	Multiple Choice	106	20	84.1%
10	Have you used a scanner or have you taken a photo/image of a text document (which may also contain images/photos)?	Yes/No	118	8	94.9%
11	Have you used an OCR software to convert paper documents into editable text files?	Yes/No	120	7	95.2%
12	For which of the following languages have you used OCR software? *	Multiple Choice	102	24	81.0%
13	How often have you used the following OCR software?	Matrix/ rating scale	114	12	90.5%
14	How capable are you with each of the following OCR Software?	Matrix/ rating scale	108	18	85.7%
15	Which of the following problems have you encountered with the OCR Software in Question 14? *	Multiple Choice	90	36	71.4%
16	Do you know that the following features are available with many OCR Software?	Matrix/ rating scale	110	16	87.3%
17	How do you rank the accuracy of OCR in the following language (If applicable)?*	Matrix/ rating scale	110	16	87.3%
18	Have you used OCR software to convert scanned handwritten text documents into an editable text document?	Yes/No	111	15	88.1%
19	If your answer to Question 18 is yes, How was the accuracy? And in which language? *	Matrix/ rating scale	69	57	54.8%
20	Please select the statement that indicates how you feel about OCR? *	Multiple Choice	106	20	84.1%
21	Comments/Suggestions/concern or anything you want to add to the survey.	Comment box	22	104	17.5%

* questions which include an “other” option

The survey questionnaire starts by collecting information on the respondent’s profile, such as gender, age, country of residence, education level, and area of expertise. The second part of the survey was divided into two parts: application software technologies and OCR software. The type of questions were mainly multiple choice, yes/no, dropdown, matrix/rating scale, and comment box. In addition, some questions had an “other” option giving the participants the flexibility to provide their own answers instead of choosing from the provided ones. The survey was conducted between January 18, 2017 and February 28, 2017. The questions used in the survey are provided in Table 2. The total number of responses were 126, approximately 15.8% of the number of emails sent, and this rate is an acceptable response rate and according to the projected expectations. During the survey period, the author sent several reminders, by email and WhatsApp to the contacts; however, this did not help much to remind people to answer the survey, and only very few responses were seen added to the number of respondents after each reminder. The low response rate could be because most

of the contacts are Doctorate and Master level individuals, and they are usually too busy to spare 10 – 15 minutes to fill the survey.

IV. RESULTS AND ANALYSIS

The survey is divided into three parts: participants personal profile information, such as gender, age, country of residence, education level and area of expertise; application software technologies; and OCR software. In this section, the questionnaire results are presented and analyzed.

A. Personal Profile Information

The total number of responses were 126, (69.4% males and 30.6% females), with two respondents who did not indicate their gender. The respondents were from 25 countries and covered the five continents, Table 3. The ages of respondents are given in Table 4. It is observed that approximately 64% of the respondents are between the ages 30 and 50, and 20% below 30 years of age. This shows that usually individuals who are employed and educated may use such software.

Table 3: The number of respondents across the continents

No.	Country	Percent Reponses	Response count	Continent	Percent Response
1	Algeria	4.0%	5	Africa	20.2%
2	Egypt	9.7%	12		
3	Ethiopia	0.8%	1		
4	Ghana	0.8%	1		
5	Morocco	0.8%	1		
6	Nigeria	0.8%	1		
7	Sudan	2.4%	3		
8	Tunisia	0.8%	1		
9	Bahrain	0.8%	1	Asia	65.3%
10	India	4.8%	6		
11	Japan	0.8%	1		
12	Jordan	2.4%	3		
13	Malaysia	8.1%	10		
14	Oman	0.8%	1		
15	Pakistan	3.2%	4		
16	Palestine	0.8%	1		
17	Qatar	0.8%	1		
18	Saudi Arabia	41.1%	51		
19	Yemen	1.6%	2	Europe	4.8%
20	Sweden	0.8%	1		
21	Ukraine	0.8%	1		
22	United Kingdom of Great Britain and Northern Ireland	3.2%	4	N. America	8.1%
23	Canada	2.4%	3		
24	United States of America	5.6%	7	S. America	1.60%
25	Brazil	1.6%	2		

*The highlighted countries are from the Middle East.

The response to question 4, “What is your area of expertise/study?” Is given in Table 5. The answer option “Other” received the following three responses: Human Resources, Information Science and Knowledge Management, and information and communications. These three responses can be added under the category Computer Science/Computer Engineering / IT, increasing the percentage of response to 60%. The education/teaching category received 14.4% followed by the sciences which received 7.2%. This shows that the main categories of respondents were from the IT, engineering and sciences sectors with very low responses from other fields. The main reason could be that the authors’ contacts were mainly from IT and engineering; however, the survey clearly mentioned that people from all areas of expertise/study are invited to fill the survey, and contacts were urged to forward the survey link to their colleagues and friends.

The level of education of respondents is very high as it can be seen from Table 6 that 88% of the respondents have post-graduate education with 57.6% Doctorate and 30.4% Master degree holders. The rest of the respondents: 8% earned or are pursuing their Bachelor degrees and 4% earned or are pursuing their High School diploma.

The respondents overwhelmingly consider that having access

to technology is “very important”, and this is very much expected given the technology era we are living in, Figure 4.

Table 4: Statistics on the age of respondents

Question 3: What is your age?		
Age category	Response Percent	Response Count
20 or younger	1.6%	2
21 – 30	14.3%	18
31 – 40	35.7%	45
41 – 50	28.6%	36
51 – 60	14.3%	18
61 – 70	4.8%	6
70 or older	0.8%	1

Table 5: Area of expertise/study of survey participants.

Question 4: What is your area of expertise/study?		
Answer Options	Response Percent	Response Count
Business and Economics	3.2%	4
Computer Science/Computer Engineering / IT	57.6%	72
Education/Teaching	14.4%	18
Engineering	3.2%	4
Government, Law, Politics and Policy	0.8%	1
Language and Linguistics	4.0%	5
Medicine, Nursing and Health Sciences	5.6%	7
Philosophy, Ethics and Theology	0.8%	1
Sciences (Biology, Chemistry, Mathematics, Physics.)	7.2%	9
Visual and Performing Arts and Film Studies	0.8%	1
Other (please specify)	2.4%	3

Table 6: Level of Education

Question 5: What is the highest level of education you have completed?		
Answer Options	Response Percent	Response Count
In High School	3.2%	4
Graduated from high school	0.8%	1
2 year Diploma	0.0%	0
1st year of university	0.8%	1
2nd year of university	0.8%	1
3rd year of university	0.0%	0
4th year of university	0.0%	0
Earned a Bachelor's degree (4 or 5 years)	6.4%	8
Earned a Master degree	30.4%	38
Earned a Doctorate degree	57.6%	72

Table 7: Statistics on how capable are individuals with application software technologies

Question 7: How capable are you with each of the following technologies to create or share documents?						
Answer Options	Very Capable Advanced	Capable Good	Somewhat Capable Acceptable	Not very Capable Poor	Never Used	Response Count
Creating text documents with a word processor	79.70%	17.10%	3.30%	0.00%	0.00%	123
Creating text documents using advanced features such as tables, images, formatting, macros... etc.	70.20%	25.00%	3.20%	1.60%	0.00%	124
Adding links to videos, audio, images into documents.	63.70%	22.60%	11.30%	0.00%	2.40%	124
Finding help from other people on the Internet using forms, message boards, social media ... etc.	33.10%	37.90%	19.40%	4.80%	4.80%	124
Programming a computer application or program for others to use	32.00%	25.40%	18.00%	12.30%	12.30%	122
Sharing or embedding video/audio files on websites.	29.50%	36.10%	20.50%	4.90%	9.00%	122
Creating and publishing a blog or online journal	23.60%	24.40%	22.00%	11.40%	18.70%	123
Creating and editing videos	21.10%	29.30%	35.80%	7.30%	6.50%	123
Creating and editing audio/sound files and recordings	21.10%	30.90%	35.00%	5.70%	7.30%	123
Designing websites using HTML, CSS, JavaScript	17.90%	26.00%	26.00%	18.70%	11.40%	123

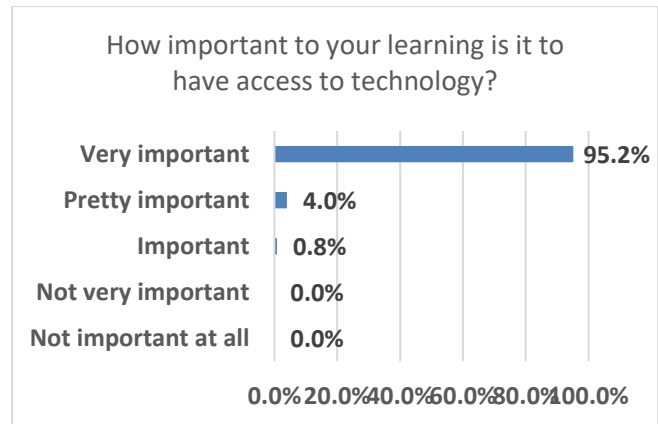


Figure 4: How important to have access to technology?

B. Application software technologies

This study investigated the capability of individuals with application software technology, Question 6, which most of us may be using on a daily basis, such as text processing, adding links to media files, creating and editing video and audio files, publishing websites, ... etc.

The results show that all participants are capable of creating basic text documents with a word processor, ranging from advanced to acceptable levels. For creating text documents, using advanced features and adding multimedia links into documents almost 98% of the participants are capable of performing such tasks. Overall, more than 70% of the participants are capable of using all the application software technologies listed in Table 7. There are some technologies that not too many people were exposed to, such as creating and publishing a blog or online journal, programming a computer application, and designing and publishing websites.

Table 8 provided the individual's usage of application software. What is interesting is that social media and word processing software are used daily by approximately 75% of the participants. Also, writing programs and using imaging software are used by approximately 15% of the participants. These software are specialized, and given the high level of education of these participants, this is expected. It is also shown

that approximately 50% of the participants use these software at least once a month. On the other hand, playing computer games, using web development software, and video processing software are rarely or never used by most of the respondents with more than 60% of the respondents answered "rarely" or "never used".

Table 8: Statistics on how often individuals use application software

Question 8: How often do you use the following application software?						
Answer Options	Daily	Weekly/ Biweekly	Monthly	Rarely	Never	Response Count
Using Social Media websites (Facebook, Twitter, ... etc)	75.80%	14.50%	3.20%	6.50%	0.00%	124
Word processing software (MS. Word, Apple iWork,... etc.)	74.80%	16.30%	4.10%	4.10%	0.80%	123
Writing programs using programming language (C++, Java, .etc.)	14.60%	14.60%	17.90%	30.10%	22.80%	123
Editing imaging software (paint, Photoshop, ... etc.)	14.40%	29.60%	22.40%	30.40%	3.20%	125
Playing computer games	8.30%	6.70%	15.80%	40.00%	29.20%	120
Web development software (HTML, CSS, Java Script, etc.)	4.90%	13.00%	21.10%	40.70%	20.30%	123
Video processing/editing software (Movie Maker, iMovie,etc.)	4.00%	9.70%	22.60%	50.80%	12.90%	124

The responses to the problems encountered when downloading application software, Table 9, show that all problems listed were encountered by at least 17% of the respondents with "the software is not free and I cannot afford buying it" receiving the highest percentage, 61.3%. Note that in, Question 9, the respondents can choose more than one answer, and those who have not encountered any problems may choose to skip the question; however, some of the respondents chose the answer option "Other" and wrote "nothing", "never used", or "no problems encountered". From the "other" responses, the respondents did not really state any problems except for one response which was "Not all software are fully localized for Middle East", and this is a real problem that many may face especially if they are not familiar with English or technical terminologies, which are very much needed during the installation process of any software.

Table 9: Problems encountered with application software

Question 9: Which of the following problems you have encountered with the application software in Question 8? (You may choose more than one answer).		
Answer Options	Response Percent	Response Count
Unable to download the software.	18.9%	20
Unable to install the software and decided to remove and not use.	17.0%	18
Had problems during installation and had to receive help online and/or from a friend.	17.0%	18
The software is not easy to use, the interface is not user-friendly.	24.5%	26
I always have to consult manuals, online help or YouTube videos (tutorials) for help.	32.1%	34
The software is not free and I cannot afford buying it.	61.3%	65
Other (please specify)	13.2%	14

C. OCR Software

In Question 10, it was asked if the individual used a scanner or have taken a photo/image of a text document which may or may not include text and images. The results show that 94.9% of the respondents have used a scanner. This was followed by Question 11, "Have you used OCR software to convert image files to editable text files?" It was interesting from the results to see that over 62% of the respondents used OCR software, either through their scanner OCR software or special OCR software. This question was followed by Question 12 on the languages used for OCR, which listed the top 10 spoken languages in the world, [42]. These are: Chinese, Spanish, English, Hindi, Arabic, Portuguese, Bengali, Russian, Japanese and Punjabi. An "other" option was also provided for the participants to include their language, if it is not on the list of languages. The "Other" option received 16 responses. The results show that the most used languages with OCR were English followed by Arabic then French, Table 10. Other languages with one or two responses include Spanish, Russian, Punjabi, Malay, and Swedish. French, Swedish and Malay were under the "Other" answer option where French received five responses, but Swedish and Malay one response each. In contrast with the rest of the responses, 9 responses of the "Other" option were "never", "not used", "nothing" ... etc. Question 13 examined the usage of the top four best OCR software in 2017, and the results are given in Table 11.

Table 10: Responses for Question 12, “For which of the following languages have you used OCR”

Question 12: For which of the following languages have you used OCR software? (You may choose more than one answer).		
Answer Options	Response Percent	Response Count
Chinese	0.0%	0
Spanish	1.0%	1
English	83.3%	85
Hindi	0.0%	0
Arabic	44.1%	45
Portuguese	2.0%	2
Bengali	0.0%	0
Russian	2.0%	2
Japanese	0.0%	0
Punjabi	1.0%	1
French	4.9%	5
Malay	1.0%	1
Swedish	1.0%	1

Table 11 shows that the best OCR software, Adobe Acrobat, has been used by approximately 84% of respondents. On the other hand, all other software have slightly, rarely or never been used by most of the respondents. A minimum of 73% of respondents never used them. Also, it is observed that most of

those who may have used these software rarely used them, and the only software with reasonable daily usage rate is Adobe Acrobat (30.7%) followed by ABBYY FineReader (1.9%). Similarly, in terms of competence in using the software, roughly the same rates for usage are reflected in the capability levels of individuals, Table 12. Overall, those who used the software daily show more competence in using the software. The survey investigated the problems which the users may have encountered during the installation process, Table 13. The results show that 41% of the respondents may need to use the software; however, they are surprised that the software is not free, and they cannot afford to buy it. In this study, 41% of the respondents consider this as a problem. In addition, 27.8% of the respondents faced difficulties, and they tend to consult technical support, manuals, tutorials, ... etc., for help in order to resolve their issues. The “Other” answer option did not receive any responses which are worth mentioning, and most of the responses were “none”, “never used”, “Arabic OCR needs some work to reach the accuracy of English” ... etc.

In addition, Question 16 presented the nine most available features in OCR software (given in Table 14) and asked the respondents if they know that such features are available in OCR software. Overall, the results showed that 56% of the users are familiar with most of the OCR available features.

Table 11: Statistics on the usage of OCR software

Question 13: How often have you used the following OCR software? and how often? (Note, Never, means you haven't heard about or used this software before?)						
Answer Options	Daily	Weekly/ Biweekly	Monthly	Rarely	Never	Response Count
Adobe Acrobat	30.7%	14.0%	8.8%	29.8%	16.7%	114
OmniPage Standard	0.9%	0.9%	4.7%	19.6%	73.8%	107
ABBYY FineReader	1.9%	2.8%	0.9%	14.2%	80.2%	106
Readiris	1.0%	1.9%	2.9%	17.1%	77.1%	105

Table 12: Statistics on the level of capability of individuals with OCR software

Question 14: How capable are you with each of the following OCR software you have used?						
Answer Options	Very Capable Advanced	Capable Good	Somewhat Capable Acceptable	Not very Capable Poor	Never Used	Response Count
Adobe Acrobat	33.3%	30.6%	15.7%	6.5%	13.9%	108
OmniPage Standard	2.9%	8.6%	8.6%	8.6%	71.4%	105
ABBYY FineReader	2.9%	5.8%	2.9%	7.7%	80.8%	104
Readiris	2.0%	8.8%	5.9%	6.9%	76.5%	102

Table 13: Problems encountered with OCR software

Question 15: Which of the following problems have you encountered with the OCR software in Question 14? (You may choose more than one answer).		
Answer Options	Response Percent	Response Count
Unable to download the software.	13.3%	12
Unable to install the software and decided to uninstall and not use.	10.0%	9
Had problems during installation and had to receive help online and/or from a friend.	13.3%	12
The software is not easy to use, the interface is not user-friendly.	7.8%	7
I always have to consult manuals, online help or YouTube videos (tutorials) for help.	27.8%	25
The software is not free and I cannot afford buying it.	41.1%	37
Please enter your statement	22.2%	20

Question 17 asks the respondents to rank the accuracy of the OCR software for the languages they have used. It was very surprising that over 91% of the respondents never used OCR for any language other than English and Arabic. Approximately 50% of the respondents ranks OCR for English as excellent, and 14.3% ranks Arabic as Excellent. However, 23.8% ranks OCR for English and 19.4% ranks OCR for Arabic as good. It is also noticed that more respondents used English OCR compared to any other language, this is followed by Arabic. The results may not be accurate and cannot be

generalized since most of the respondents were from the Middle East, and very few people represented other languages mentioned in Table 15.

To further explore the use of OCR software, Question 18 asked if the respondents used OCR software to convert handwritten text documents into editable text documents. The results show that about a third of the respondents (32.4%) used OCR for handwritten documents. Also, Question 19 presents the OCR software accuracy level for handwritten documents.

Table 14: OCR Features

Question 16: Do you know that the following features are available with many OCR software?			
Answer Options	Yes	No	Response Count
Multi language support	72.2%	27.8%	108
The output file retains layout of original scanned paper document	63.2%	36.8%	106
The output file retains Fonts of original scanned paper document.	51.9%	48.1%	106
The output file retains Tables from original scanned paper document	52.8%	47.2%	106
OCR software can de-skew (rotate) image	50.9%	49.1%	106
Multi-page document recognition	58.1%	41.9%	105
Integrates with Cloud Storage.	41.0%	59.0%	105
Integrates with MS Office applications.	66.4%	33.6%	107
Integrates with HTML.	42.5%	57.5%	106
Average	56%	44%	106

Table 16, provides the results for Question 19, presenting the estimated accuracy from the user's perspective for the languages used by the respondents. The results show that a minimum of 93% of respondents never used OCR for handwritten documents. The accuracy level, as expected, is very low compared to printed documents. Here, English received the highest accuracy response rate of 18.8% as "Excellent" i.e., above 85% accuracy, followed by Arabic, 10.6%.

To conclude, the respondents were asked to provide the statement/s on how they feel about OCR. Table 17 shows the statement, "OCR is a very important technology and must be available as a feature in all text/image processing software," receives the highest response with 78.3%, followed by the statement, "OCR is very important to be used by universities and public libraries to convert old documents to digital documents," with 41.5%.

Table 15: Ranking the accuracy of OCR software for different languages using printed documents

Question 17: How do you rank the accuracy of OCR in the following languages? if applicable?						
Answer Option	Excellent (Above 85%)	Good (70% - 84.9%)	Acceptable (50% - 69.9%)	Poor (Below 50%)	Never Used	Response Count
Chinese	0.0%	3.6%	2.4%	1.2%	92.8%	83
Spanish	0.0%	3.8%	2.5%	0.0%	93.8%	80
English	49.5%	23.8%	9.5%	4.8%	12.4%	105
Hindi	1.2%	3.7%	2.4%	2.4%	90.2%	82
Arabic	14.3%	19.4%	16.3%	11.2%	38.8%	98
Portuguese	0.0%	6.0%	1.2%	0.0%	92.9%	84
Bengali	0.0%	1.2%	2.4%	2.4%	93.9%	82
Russian	0.0%	2.5%	4.9%	1.2%	91.4%	81
Japanese	0.0%	3.7%	1.2%	2.4%	92.7%	82
Punjabi	0.0%	1.2%	3.7%	2.5%	92.6%	81

Table 16: Ranking the accuracy of OCR software for different languages using handwritten documents

Question 19: If your answer to Question 21 is yes, How was the accuracy? and in which language?						
Answer Option	Excellent (Above 85%)	Good (70%-4.9%)	Acceptable (50% - 69.9%)	Poor (Below 50%)	Never Used	Response Count
Chinese	0.0%	3.5%	0.0%	0.0%	96.5%	57
Spanish	0.0%	1.8%	0.0%	1.8%	96.5%	57
English	18.8%	17.4%	14.5%	4.3%	44.9%	69
Hindi	1.8%	0.0%	0.0%	5.4%	92.9%	56
Arabic	10.6%	9.1%	12.1%	6.1%	62.1%	66
Portuguese	0.0%	1.8%	0.0%	0.0%	98.2%	57
Bengali	1.8%	0.0%	0.0%	1.8%	96.5%	57
Russian	0.0%	1.8%	0.0%	1.8%	96.5%	57
Japanese	0.0%	1.8%	0.0%	1.8%	96.5%	57
Punjabi	0.0%	0.0%	1.8%	0.0%	98.2%	57

Table 17: Responses to “Select the statement that indicates how you feel about OCR”

Question 20: Please select the statement that indicates how you feel about OCR? (You may choose more than one statement).		
Answer Options	Response Percent	Response Count
OCR is a very important technology and must be available as a feature in all text/image processing software.	78.3%	83
OCR is a technology that is not mature enough to be used for some languages which written from right to left, such as Arabic.	21.7%	23
OCR is a technology that is not mature enough to be used for many languages.	16.0%	17
OCR is acceptable to be used for printed text; however, with handwritten text it is preferred not to be used.	19.8%	21
OCR is very important to be used in government offices to convert handwritten documents to editable files.	22.6%	24
OCR is very important to be used by universities and public libraries to convert old documents to digital documents.	41.5%	44
OCR is a saver for old documents and manuscripts from being lost/stolen/burned etc.	36.8%	39
Other (please specify)	8.5%	9

Finally, the respondents were given the chance to provide comments and/or suggestions about the survey. From the responses, almost all of the comments were “thank you”, “nothing”, and “good luck” comments except for the following three interesting comments: the first comment was “Thanks for providing me an opportunity to be part of your research and share my contribution. I think that expertise of university teacher’s use of software and its related knowledge depends upon the availability of software in the institution. If an institution has enough sufficient software packages, then its next responsibility is to train its employees in using that software. So, institutions should have latest software and expertise in using and providing training in use of that software.” The second comment was “I am not familiar with most of the software, and I think it is useless to have these many software”. The third comment is, “I only use the software provided by the operating system, and those available on my computer at work.” These comments summarize the reasons why people are not familiar with many software. We can say that we spend most of our time at work, and this influence the type of software we may use depending on our areas of expertise.

V. LIMITATIONS AND DISCUSSION

In general, document analysis software is very essential. Hence, OCR reduces time for processing data collection, which

if done manually (data entry), takes longer time and is prone to human errors. The study findings may only benefit very limited number of people who are interested in English/Arabic OCR and image editing software and, therefore, not too many people participated in this survey. In addition, the low response rate could be because there are no monetary incentives for filling the survey. The results show that people were overwhelmed with the number of available software which most of them never heard about. Even though the Top four OCR software were listed, the results show that approximately 73% of the respondents never used or even heard about these software.

There are several limitations in the study findings. The study results cannot be generalized since the data collection was not purely random collected from faculty members in the IT sector to widen the sampling group to include individuals from all different sectors. The study reported approximately 58% of respondents are from the IT sector. The survey was distributed online to people who are familiar with using the Internet, sending emails, and using social media. The data collection was geographically distributed all over the world; however, 58% of the respondents were from the Middle East and about 71% of the Middle East respondents were from Saudi Arabia.

The accuracy rate reported by the commercial OCR software does not reflect the results obtained in this study. For example, the study reported the recognition rate for English, Latin script, as excellent with only 50% of the respondents, which is still not

100% accuracy as reported in [40]. Therefore, further investigation of the accuracy rates is needed for all language scripts including Latin scripts.

In the future, this study needs to be distributed to a wider range of people especially young people who are technology talented, and this survey did not reach them. In this study, 20% of the respondents were below 30 years old with only 1.6% were 20 or younger. Although this study provides insights into the familiarity of individuals with document and OCR technology corresponding to the age and gender, future research support is needed to understand the familiarity and capability of individuals with different software based on different age groups and gender.

The area of expertise and employment place of an individual pretty much have an influence on most of the types of software used in addition to the influence of the work environment which also influences the type of social media used. Another important finding from the survey is that technical support for the software in multi-languages, especially Arabic, is not available, which may be a factor that such software is not popular to use.

Currently, companies are providing free or limited usage of online OCR service or OCR mobile apps. This provides a real-time processing which can be used anytime from a personal computer or a mobile device. Future study should also evaluate this service and its possible replacement for current software installed on desktop computers.

A framework for testing OCR software is urgently needed to be able to test multilingual document analysis and OCR software. Databases of documents to test this software for all languages need to be developed in a similar fashion to be able to test different languages and provide comparable results for any language under any OCR platform. Therefore, the author recommends that more research is encouraged for different languages other than Latin, where the recognition rates are far below from those reported for Latin scripts, especially cursive scripts and printed handwritten scripts for all languages including Latin.

VI. CONCLUSION AND FUTURE WORK

The OCR technology went through a major progress has progressed a great amount since it started in the early years of the last century. The study has confirmed that document analysis software and OCR, in particular, are very important and essential technology, and its use is seen in many applications around us. The recognition of cursive text in any language is still an active area of research. Higher recognition rates for printed or handwritten text is not possible without the use of contextual or grammatical information. The work carried out in this paper is the first of its kind and it sheds some light on the user perspective and their familiarity with document analysis software, especially OCR. The study concludes that a framework for testing OCR software is urgently needed to be able to test multilingual document analysis and OCR software. Finally, more research is encouraged for different languages other than Latin scripts.

ACKNOWLEDGMENTS

The author would like to thank all colleagues and their friends who voluntarily filled the questionnaire survey for this study. In addition, the authors would like to thank the anonymous reviewers for their valuable comments and necessary suggestions that helped to clarify and improve the quality of the paper.

REFERENCES

- [1] B. B. Chaudhuri, (Ed.), "Digital document processing: major directions and recent advances," *Springer Science & Business Media*, 2007.
- [2] H. D. Wactlar, M. G. Christel, Y. Gong and A. G. Hauptmann, "Lessons learned from building a terabyte digital video library," *Computer*, vol. 32, no. 2, pp. 66-73, Feb, 1999.
- [3] H. F. Herbert, "The history of OCR, optical character recognition." *Manchester Center, VT: Recognition Technologies Users Association*, 1982, ISBN 9780943072012.
- [4] R. White, and T. Downs, "How computers work," *Que Corp.*, 2007.
- [5] S. Singh, "Optical character recognition techniques: a survey," *Journal of emerging Trends in Computing and information Sciences*, Vol. 4, no. 6, pp. 545-550, 2013.
- [6] <https://ocr.space/ocrapi>, <Accessed 20 December 2016>.
- [7] S. Naz, A.I. Umar, S. H. Shirazi, S. B. Ahmed, M. I. Razzak, and I. Siddiqi, "Segmentation techniques for recognition of Arabic-like scripts: A comprehensive survey," *Education and Information Technologies*. Vol. 21, no. 5, pp. 1225-1241, 2016.
- [8] Y. M. Alginahi, "Computer analysis of composite documents with non-uniform background," Thesis, University of Windsor, 2004.
- [9] Y. M. Alginahi, D. Fekri, and M.A. Sid-Ahmed, "A neural-based page segmentation system," *Journal of Circuits, Systems, and Computers*, 14(01), pp.109-122, 2005.
- [10] Y. M. Alginahi, M. Mudassar, and M. N. Kabir, "An Arabic Script Recognition System." *KSIIT Transactions on Internet and Information Systems*, Vol. 9, no. 9, 3701-3720, 2015.
- [11] I. Kaur and S. Mahajan, "Bilingual Script Identification of Printed Text Image," *International Journal of Engineering and Technology*, Vol. 2, no. 3, pp. 768-773, 2015.
- [12] Y. M. Alginahi, "A survey on Arabic character segmentation," *International Journal on Document Analysis and Recognition (IJ DAR)*, Vol. 16, no. 2, pp.105-126. 2013.
- [13] N. Fareen, M. A. Khan, and A. Durrani, "Survey of Urdu OCR: an offline approach," *Proceedings of the Conference on Language & Technology*, pp. 67-72, 2012.
- [14] Parvez, Mohammad Tanvir, and Sabri A. Mahmoud. "Offline Arabic handwritten text recognition: a survey," *ACM Computing Surveys (CSUR)*, Vol. 45, no. 2, 2013.
- [15] M. Shatnawi, Off-line handwritten Arabic character recognition: a survey. *International conference on image processing, computer vision, and pattern recognition (IPCV)*, pp. 52 – 58, 2015.
- [16] U. Saeed, "Automatic Recognition of Handwritten Arabic Text: A Survey." *Life Science Journal*, Vol. 11, no. 3s, 2014.
- [17] D. N. Hakro, A. Z. Talib, Z. Bhatti, and G. N. Moja, "A Study of Sindhi Related and Arabic Script Adapted languages Recognition," *Sindh University Research Journal-SURJ (Science Series)*, Vol. 46, no. 3, pp. 323-334, 2014.
- [18] D. N. Hakro, A. Z. Talib, Z., and G. N. Moja, "Multilingual Text Image Database for OCR," *Sindh University Research Journal-SURJ (Science Series)*, Vol. 47, no. 1, 2016.
- [19] D. N. Hakro, and A. Z. Talib, "Printed Text Image Database for Sindhi OCR," *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 15, no. 4, 2016.
- [20] J. Sadri, M. R. Yeganehzad, and J. Saghi, "A novel comprehensive database for offline Persian handwriting recognition," *Pattern Recognition*, Vol. 60, pp. 378-393, 2016.
- [21] M. Wahab, H. Amin, and F. Ahmed, "Shape analysis of Pashto script and creation of image database for OCR," *IEEE International Conference on Emerging Technologies (ICET 2009)*, pp. 287-290, 2009.
- [22] M. Sagheer, C. He, N. Nobile, and C. Suen. "A new large Urdu database for off-line handwriting recognition," *International Conference on Image Analysis and Processing*, pp. 538-546. Springer Berlin Heidelberg, 2009.

- [23] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri. "IFN/ENIT-database of handwritten Arabic words," *Francophone International Conference on writing and Document*, CIFED'02, Vol. 2, pp. 127-136, 2002.
- [24] S. Mir, S. Zaman, and M. W. Anwar. "Printed Urdu Nastalique Script Recognition Using Analytical Approach." *IEEE 13th International Conference on Frontiers of Information Technology (FIT)*, pp. 334-340, 2015.
- [25] R. Ahmad, M. Z. Afzal, S. F. Rashid, and S. Naz, "Semi-Automated Transcription Generation for Pashto Cursive Script," *J. Appl. Environ. Biol. Sci.*, Vol. 6, no. 3S, pp. 96-101, 2016.
- [26] S. Nasrollahi, and A. Ebrahimi, "Printed Persian Subword Recognition Using Wavelet Packet Descriptors," *Journal of Engineering*, 2013.
- [27] D. V. Kaur, "Comparison Analysis of Post-Processing Method for Punjabi Font." *International Research Journal of Engineering and Technology (IRJET)*, Vol. 4, no. 1, pp. 1206 – 1211, January 2017.
- [28] Q. Chen, B. S. Yuan, X. LI, H. Y. REN, and J. H. Zhang, "Printed Uyghur Character Recognition Based on Template Matching," *Computer Technology and Development*, Vol. 4, 2012.
- [29] A. A. Sanjrani, J. Junaid, M. Bakhtyar, W. Noor, and M. Khalid, "Handwritten Optical Character Recognition system for Sindhi numerals," *International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, pp. 262-267, 2016.
- [30] Z. Jan, M. Shabir, M. A. Khan, A. Ali, and M. Muzammal, "Online Urdu Handwriting Recognition System Using Geometric Invariant Features," *Nucleus*, Vol. 53, no. 2, pp. 89-98, 2016.
- [31] Ahmad, Riaz, Muhammad Zeshan Afzal, Sheikh Faisal Rashid, Marcus Liwicki, Andreas Dengel, and Thomas Breuel. "Recognizable units in Pashto language for OCR." *13th International Conference on Document Analysis and Recognition (ICDAR '15)*, pp. 1246-1250, 2015.
- [32] M. Kazemi, M. Yousefnezhad, and S. Nourian. "A New Approach in Persian Handwritten Letters Recognition Using Error Correcting Output Coding," *Journal of Advances in Computer Research*, Vol. 6, no. 4, pp. 107-124, 2015.
- [33] G. Z. Campus, G. N. NIET, and P. Bathinda, "Review: A Literature Survey on Text Segmentation in Handwritten Punjabi Documents," *Research Cell: An International Journal of Engineering Sciences*, Vol. 3, pp. 174 – 178, December 2014.
- [34] H.Y. Ren, B.S. Yuan, and Y. Tian, "On-line & Handwritten Uyghur Character Recognition Based on BP Neural Network," *Journal of Microelectronics & Computer*, Vol. 8, p.060, 2010.
- [35] D. N. Hakro, I. A. Ismaili, A. Z. Talib, Z. Bhatti, G. N. Mojai, "Issues and challenges in Sindhi OCR," *Sindh University Research Journal-SURJ (Science Series)* Vol. 46, no. 2, 2014.
- [36] E. Borovikov and I. Zavorin. "A multi-stage approach to Arabic document analysis. Guide to OCR for Arabic scripts," 2012, *Springer*, London, pp. 55-78.
- [37] R. Holley, "How good can it get? Analysing and improving OCR accuracy in large scale historic newspaper digitization programs." *D-Lib Magazine*, Vol. 15, no. 3/4, 2009.
- [38] P. R. Dussawar and P. B. Jha. "Text Extraction from Complex Color Images Using Optical Character Recognition," *International Journal of Science and Research (IJSR)*, Vol. 4, no. 7, pp. 730 – 735, July 2015.
- [39] C. Y. Suen, R. Plamondon, A. Tappert, A. J. W. M. Thomassen, J. R. Ward, and K. Yamamoto. "Future challenges in handwriting and computer applications." *Third International Symposium on Handwriting and Computer Applications, Montreal*. May 1987.
- [40] <http://www.toptenreviews.com/business/software/best-ocr-software/>, <Accessed 12 November, 2016 >
- [41] <http://download.cnet.com/s/ocr/>, <Accessed 15 November, 2016 >
- [42] <https://www.babble.com/en/magazine/the-10-most-spoken-languages-in-the-world>, <Accessed 29 November, 2016 >



Yasser M. Alginahi, earned a Ph.D., in Electrical Engineering from the University of Windsor, Ontario, Canada, and a Master of Science in Electrical Engineering from Wright State University, Ohio, U.S.A. He is a licensed Professional Engineer, Ontario, Canada, a member of Professional Engineers Ontario, a senior member of IEEE and IACSIT since 2010. Dr. Alginahi is an Associate Professor at the Department of Computer Science, Deanship of Academic Services, Taibah University. He is also the consultation unit coordinator at the IT Research Center for the Holy Quran and its Sciences, Taibah University. Dr. Alginahi is on editorial board of several international journals. He published a book entitled "Document Image Analysis" and he has published over 80 journal and conference papers. He worked as Principle Investigator and co-Principle Investigator on many funded research projects by the Deanship of Scientific Research at Taibah University and other organizations such as King Abdul-Aziz City of Science and Technology. His current research interests are Information Security, Document Image Analysis, Pattern Recognition, OCR, Modeling and Simulation, Intelligent Transportation Systems and Numerical Computations.

Digital Holy Quran Authentication, Certification and Standardization: A Survey

Yasser M. Alginahi^{1,2}, Muhammad Khurram Khan³, Omar Tayan^{2,4}, Mohammed Zakariah⁵,

¹Dept. of Computer Science, Deanship of Academic Services, Taibah University, KSA

²IT Research Center for the Holy Quran and Its Sciences (NOOR), Taibah University, KSA

³Center of Excellence in Information Assurance (CoEIA), King Saud University, KSA

⁴College of Computer Science and Engineering, Dept. of Computer Engineering, Taibah University, KSA

⁵Research Center of College of Computer and Information Sciences, King Saud University, KSA

yasser.m.alginahi@ieee.org, mkhurram@ksu.edu.sa, otayan@taibahu.edu.sa, mzakariah@ksu.edu.sa

Abstract- The digital authentication, certification and standardization of Information Technology (IT) applications of the Holy Quran has attracted researchers and organizations to explore its importance and requirements. This is an emerging area of research that needs more attention since it investigates sensitive digital documents and applications. The objective of this study is to gather information from users, developers, designers, decision makers and organizations on their perception on authentication, certification and standardization of IT applications for the Holy Quran. This study is based on a questionnaire survey which was distributed online to highly education individuals from all over the world. The study emphasized the need for more security and quality assurance in IT Quran applications. In addition, standardization and certification through authentic Quran organizations are encouraged to validate the Quranic content. The study results and analysis were extracted from over 80 questions divided between six categories of participants. The survey received 500 responses from 31 countries and five continents. The findings of the study are considered a step towards the importance of secure sensitive documents in all different areas of research particularly digital Quran content.

Keywords- Quran, Authentication, Security, Quality of Service, Standards

I. INTRODUCTION

Quran is the sacred book of Islam which was revealed to Prophet Mohammed (peace be upon him) over 14 centuries ago. Quran is usually read from paperback format called Mushaf; however, nowadays the recent advancements in technology allowed the use of smart digital gadgets which are portable and adopted by almost everyone [1]. Therefore, the rapid advancements in technology with the worldwide spread of the Internet allowed for the dissemination of online digital multimedia content. The increase in speed and storage capabilities allowed for a rapid increase in the digitization of all sorts of multimedia publications especially the digitization of printed documents has seen an exponential jump as archives are digitized and offices are eliminating papers and becoming paper free. Therefore, Quranic organizations and research centers exploited this movement of digitization to benefit in the development of Quranic Web and mobile apps. In this section, the most recent research work related to Quran computing i.e., security, authentication and standardization is presented.

Due to the sensitivity of Holy Quran, it is very crucial to authenticate the verses or parts of the Quran available through webpages to avoid intentional and unintentional distortion/forgery to Quranic verses. The goal is to use information retrieval techniques to check for documents that include unauthentic Quranic

verses, therefore, authentication techniques are required to check the world-wide web for any unauthentic Quranic verses. Thus, this is possible but false results may appear if other different Quranic readings text which may be found in some websites [8].

The work by Tayan and Alginahi, conducted a survey to investigate the significance of online and offline software applications and their benefits and effectiveness for users. The study showed that approximately 80% of participants had used technology in Quran memorization, technology is more appealing to the younger generation and 50% of the participants preferred Internet/software or handheld/portable technologies as compared with traditional methods. The study concluded with several challenges from the user-perspective these are: English-language barrier since many applications are in English, Technology illiteracy, limited Internet resources and connectivity. In addition, the participants provided suggestions for new applications and modifications to existing technology [2]. In another work, Tayan and Alginahi presented a review of different multimedia watermarking techniques which are applicable to sensitive digital content and discusses how those approaches can accomplish the required protection [3].

The work in [4] presented a review on the types of Information Security (IS) aspects which include; data-storage, in-transit data and data access prevention for unauthorized users. The work presented the main challenges in IS with regards to vulnerabilities and data-breaches in addition to discussing the mechanisms for enhancing data protection, IS policies and standards for protecting data content [4].

The survey in [5] provided a review on recent researches on Digital Holy Quran authentication, protection and integrity authenticity, it focused on analyzing and categorizing the existing research related to preserving and verifying the content integrity of the Holy Quran. It concluded by a recommendation to develop a reliable universal database of authentic and verified Digital Quran and hadith content, in addition, to developing a Real-Time Quran Verse Detection Expert System with improved accuracy and precision [5].

The research paper “Digital Quran Computing: Review, Classification, and Trend Analysis” by Zakariah et al. provided a subject/theme categorization of digital Quran research based on the topical trends then a discussion of their key features, limitations, and research directions. Next, a set of recommendations regarding security and authentication, standardization and quality of service, unified translation, and E-Learning Approaches and Quran Knowledge base were presented. Finally, the authors concluded by discussing the open challenges in digital Quran Computing [6].

In [7], Sabbah and Selamat proposed a framework authenticity detection method for text Quranic verses extracted from online forum posts based on computing numerical Identifiers of words in the detected text then comparing these identifiers with Identifiers of original Quranic manuscript. The results show that the average accuracy rate was 62%, precision 75% and recall 78%. The authors stated that their future work will incorporate computational intelligence methods to increase the authentication which will also involve other Quran media such as sound, images and video to improve the detection.

Alsmadi and Zarour designed a tool to evaluate the integrity of the wording in the e-versions of the Quran by generating the metadata related to all words in the Quran i.e., to preserve the counts and

locations. This is like hash algorithms used in checking the integrity of data files in a disk, as a result a tiny modification will result in a different hash function. The results show that hashing verification can be a good candidate for the automatic Quran authentication process with high confidence [8].

The exponential increase in the number of websites on the Internet the vulnerability of sensitive information, such as digital Quran, maybe at stake. From the huge number of Islamic websites not many of them present digital copies of the Quran that are verified and certified by an authentic source. The work by Mostafa and Ibrahim propose a computer-based system on the public key infrastructure and the digital signature to secure and verify the content of Holy Quran script on the web [9]. The authors claim its applicability to the application of the Holy Quran.

Digital copies of fake or distorted Quran have been detected online and in some Quran mobile apps. Therefore, to eliminate such forgery, by authentication, which may be intentional or unintentional will solve most the threats and provide confidence in users to use such web and mobile applications. The study in [10] presents a comprehensive research survey of works conducted in the area of Quran authentication from an information security perspective. This study concluded the urgent need to provide a good content security, and integrity of digital Quran [10].

Cryptography have been used to secure data against tampering attempts and protects the data integrity. The work by AlAhmad, Alshaikhli, and Jumaah proposed a cryptography algorithm (Combination between AES and RSA Cryptography Algorithms (CARCA)). The results show that the CARCA method, with the two encryption algorithms, showed a boost and improvement in the protection of the Digital Holy Quran Hash Digest [11].

The works in [12 - 13] presented a watermarking method to enable the authentication and detection of the image forgery on the digital Quran images. The proposed method uses two layers of embedding scheme. First, the discrete wavelet transforms are applied to decompose the host image into wavelet prior to embedding the watermark in the wavelet domain. Then, the watermarked wavelet coefficient is inverted back to the spatial domain where the least significant bits are utilized to hide another watermark. Next, a chaotic map is used to blur the watermark to secure it against the local attack. This technique provides high watermark payloads, though preserving the image quality.

On the issue of certification, Khan, Siddiqui and Tayan presented a digital Quran certification framework by utilizing modern digital authentication and certification techniques i.e., a certification authority and religious scholars follow a rigorous procedure that checks the requirements process for certification and upon approval a digital certificate is issued for the application. This framework controls the digital content of the Holy Quran and doesn't allow any modifications by the users [14].

To the best knowledge of the authors the literature provides a good quality of security and authentication, however, very scarce work on Quran certification and no work available on standardization of digital Quran. As presented in this section most of the available Quran authentication techniques have been used for the purpose of research and no commercial implementation of such techniques is provided. On the issue of certification digital copies of the Quran what is available is a framework for certification. On the

other hand, the issue of standardization has not been address in literature. Therefore, the available research work is considered a positive step for researchers in this area to come together and set procedures and standards for security and authentication, certification and standardization of digital Quran.

In this paper, the security and authentication, certification, and standardization challenges regarding the development of Quran apps are presented from the perspective of users, developers, designers, decision makers and institutions. The study was based on a survey conducted to gather information on technology adoption for reading the Quran, security and quality control issues in development of Quran software (mobile and web applications), in addition to standardization issues of institutions and software. This work is unique as it gathered feedback/perspective from six different categories to identify all aspects of authentication, certification and standardization from the perspective of the users, developers and programmers, designers, managers, organizations and publishers.

II. METHODOLOGY

A semi-structured questionnaire survey was developed using monkeysurvey.ca and distributed online to the authors contacts who are mainly highly educated individuals from all over the world. The survey was conducted to discover This survey is delivered and distributed using a range of social media tools, and its findings are taken to characterize sampling of English/Arabic speaking users of Quran products. The main objective from this study is gathering the perspective/feedback of different categories of individuals and organization on the requirements and importance of digital authentication, Certification and Standardization of IT applications for the Holy Quran. Before the survey was conducted it was piloted to ten subjects who were interviewed so that to provide their feedback on the questions of the survey and comment on its clarity. Next, the semi-structured survey was designed then distributed online to over 3500 emails and social media accounts including WhatsApp. The response rate of the survey was approximately 14%. The survey was distributed in both English and Arabic. The total number of respondents was 500, from whom 307 filled the English survey and 193 filled the Arabic survey. The survey was designed using monkeysurvey.ca, the questions types include, Yes/No, Multiple choice Questions, and open-ended questions.

TABLE 1
THE PERSONAL PROFILE QUESTIONS ASKED

No.	Question	Type of Question	Completed Responses		Skipped Responses	
			Arabic	English	Arabic	English
1	What is your gender?	Multiple Choice	184	300	9	7
2	Which age group do you belong to?	Multiple Choice	184	302	9	5
3	What is your highest qualification?	Multiple Choice	184	301	9	6
4	Your employment status?	Multiple Choice	181	302	12	5
5	What is your Nationality?	Dropdown	173	292	20	15
6	Which category/occupation do you belong to?	Multiple Choice	185	289	8	18

The survey is divided into two parts, the first part includes personal questions and the second part includes specific questions targeting certain categories of individuals and organizations by requesting them to provide their feedback on questions related to requirements and importance of digital authentication,

certification, standardization of IT applications for the Holy Quran. The survey with questions on the personal profile of the respondents, including, age, gender, nationality, employment status and education level. Table 2, shows that the highest response rate for this survey was from the age group 35 – 44 years old with approximately 41%, followed by 28% for the age group 45 – 54, then 19.8% for the age group 45 – 55 and 8.2% for the age group 55 - 64. It is noticed that the age groups below 25 and above 65 counted for about 3% of the respondents.

TABLE 2
AGE GROUPS OF PARTICIPANTS

Answer Options	Response Percent	Response Count
15-24	2.1%	10
25-34	19.8%	96
35-44	40.9%	199
45-54	28.0%	136
55-64	8.2%	40
65+	1.0%	5

From both surveys (Arabic/English) conducted, it was noticed that 44.6% of the participants are females and 55.4% are males. Table 3 shows that the minimum completed education level was a high school diploma and approximately 89.3% of the participants have a bachelor or graduate degree with approximately 51% hold a Doctorate degree which shows that most of the participants are highly educated.

TABLE 3
THE EDUCATION LEVEL OF PARTICIPANTS

Answer Options	Response Percent	Response Count
High School	2.9%	14
Associate/Vocational/Trade Degree	0.2%	1
Certified Professional/Diploma Holder	0.2%	1
Bachelor Degree Holder	5.2%	25
Master Degree Holder	26.2%	127
Doctorate Degree Holder	50.9%	247
Postdoctoral	12.2%	59
Prefer not to say	2.3%	11

The employment status of the participants shows that 91% are employed full-time, part-time or self-employed, 6% are students, 0.4% are retired, 1.4% unemployed and 1.2% preferred not to say. The nationalities of the participants included 31 countries from five continents. The following are the countries of the participants: United States (3), Algeria (38), Australia (4), Bahrain (3), Bangladesh (2), Canada (7), China (1), Egypt (34), Ethiopia (1), India (33), Indonesia (4), Iran (3), Iraq (7), Jordan (65), Kuwait (6), Lebanon (3), Libya (4), Malaysia (78), Morocco (9), Nigeria (3), Oman (7), Pakistan (54), Palestine (10), Saudi Arabia (23), Sudan (18), Spain (1), Syria (10), Tunisia (4), Turkey (3), United Kingdom (10), and Yemen (17). The number between parenthesis indicates the number of participants from each country.

After responding to the personal profile section of the survey, the participants have to choose the category they belong to from the six categories shown in Table 4 before moving to the second part of the survey and answer the corresponding questions related to that category. Table 4, provides the categories and the number of respondents in each category.

TABLE 4
THE CATEGORIES OF PARTICIPANTS

Name of Category	Response Percent	Response Count
1- A general IT/Software User or Advanced User	47.3%	224
2- Software/Mobile/Digital Application Developer	18.4%	87
3- Graphics Designer	1.9%	9
4- Organization Manager/Director/Decision Makers/Owners/CEOs	22.8%	108
5- Islamic Institutes	6.8%	32
6- Holy Quran Publishers	3.0%	14

Even though Table 4 shows that 95% of the 500 participants completed the personal profile by choosing a category only 350 participants, i.e., 70% of the total no. of participants, completed the questions of the category they chose in Table 4. Finally, Table 5 provides the responses for each category. The results show that 51.4% are general IT/software users, 18.3% are software developers, 22.3 managers and decision makers, with less than 7% Islamic institutes and Holy Quran publishers.

TABLE 5
THE RESPONSE RATE OF PARTICIPANTS FOR EACH CATEGORY

Name of Category	Response Percent	Response Count
1- A general IT/Software User or Advanced User (15 Questions)	51.4%	180
2- Software/Mobile/Digital Application Developer (16 Questions)	18.3%	64
3- Graphics Designer (8 Questions)	1.1%	4
4- Organization Manager/Director/Decision Makers/Owners/CEOs (15 Questions)	22.3%	78
5- Islamic Institutes (9 Questions)	5.7%	20
6- Holy Quran Publishers (11 Questions)	1.1%	4

III. RESULTS ANALYSIS & DISCUSSION

This section presents the analysis and discussion for the results from the six survey categories.

A. General IT/Software User or Advanced User

The list of questions for this category is provided in Table 6.

TABLE 6
THE GENERAL IT/SOFTWARE USER OR ADVANCED USER CATEGORY QUESTIONS

No.	Question	Response Count
1	How often do you recite Qur'an?	180
2	Do you also recite Qur'an on-line?	180
3	What is the reason for not reciting the Qur'an on-line? (if the previous answer is NO)	69
4	Is your level of confidence in digital Quran authenticity equal to when compared with printed Quran books?	180
5	Do you use Digital Devices to read Qur'an (Smartphones/Tablets/ digital diary etc.)?	181
6	Why you do not recite the Qur'an on a digital device? (if the above is NO)*	47
7	In your opinion, are the current digital copies of the Quran available on different digital devices authentic?	179
8	In your case, do you prefer a digitally signed and 100% authentic copy of the Quran on-line or on digital device?	180
9	Did you ever encounter a fake copy of Qur'an available on-line or on a digital device?* Please provide URL.	177
10	What are the main facilities/services you use from online, web-based or smartphone Quran apps?*	186
11	Do you think it is necessary for a Quran Authentication Body to monitor and endorse the digital copies of the Qur'an Worldwide?	181
12	Are you pleased with the features of most online/smartphone Quran apps?	176
13	Please state the name of your favorite Quran Application smartphone or URL:	72
14	Are you pleased with the quality of information provided by most apps?	175
15	What is your vision/comments/recommendations of how new technologies could be used to develop and improve Quranic/Related Smartphone Apps, Web Apps or user interaction further?	79

*Question has an "Other" option or "comment box"

The results show the users response to questions related to their personal use and confidence in digital Quran, in addition to their perception on authenticity. Table 7 provides the response of the participants to the question “How often do you recite Quran?” It shows that over 65% recite the Quran daily or on alternative days and 14.4% recite it weekly which means that at least 80% of respondents reads Quran at least once a week. Also, 62.2% recite Quran online besides reading from a paper copy of the Quran, and 37.8% only prefer reading from paper copies of the Quran. From those who do not read Quran online, 92.8% feel uncomfortable reading from a computer or smart mobile screen and 7.2% specified that they are not Internet users. However, the response to “Do you use digital devices (Smartphones/Tablets/Digital diary ...etc.) to read Quran?” shows that 82.3% use digital devices compared to 17.7% who do not use any gadgets to read Quran. The reasons for not reading from digital devices include: medical reasons, feeling uncomfortable, feeling uncertainty of the authenticity of the Quran, prefer to listen from Musha’af (paper copy of the Quran), and getting more benefits from reading from Musha’af. These results emphasize the importance of reading the Quran regularly and the need to provide ways to make it easier to recite and memorize it.

TABLE 7
RESPONSE OF THE PARTICIPANTS TO THE QUESTION “HOW OFTEN DO YOU RECITE QURAN?”

Answer Options	Response Percent	Response Count
I have memorized the Qur'an by heart	2.2%	4
Daily	46.7%	84
Alternate days	16.7%	30
Once/Week	14.4%	26
Once/Month	4.4%	8
Once/Year	1.7%	3
Rarely	5.6%	10
Prefer not to say	8.3%	15
Doesn't apply to me	0.0%	0

The level of confidence in digital Quran authenticity compared with printed Quran paper back could be another reason why people do not use or trust digital copies of the Quran. The results show that 41.6% of the respondents never had any peculiar feelings at all in regard to the level of confidence in digital Quran authenticity, 20.6% always feel that the contents may be produced by an un-authentic source, 17.8% always have a feeling that the contents may be forged/modified or not properly scrutinized and 21.1% never thought about it.

Responding to the question “In your opinion, are the current digital copies of the Quran available on different digital devices authentic?” the response shows that 21.2% believe that the digital copies of Quran are authentic and 7.3% think otherwise. However, 60.3% are not sure and 11.2% never thought about it. This shows that many people assume what they download from app stores is authentic and such an issue does not cross their minds. In addition, most of the people do not memorize the Quran and may not even notice if there is a mistake. Therefore, the next question “do you prefer a digitally signed and 100% authentic copy of the Quran on-line or on digital device?” which addresses authenticity shows that 85% think it is very essential and will give authenticity a very high preference, 12.8% agree and may prefer 100% authenticity and 2.2% do not prefer digitally signed copy of the Quran. In response to the Question

“Did you ever encounter a fake copy of Qur’an available on-line or on a digital device?” 12.4% mentioned that they encountered a fake copy of the Quran. The True Furqan is one main example provided by participants, some asked to refer to google images and others could not recall the names of the websites that they encountered a copy of fake or modified Quran. This raises the question “Is it necessary for a Quran Authentication Body to monitor and endorse the digital copies of the Qur’an Worldwide?” and as expected the response of the participants highly emphasize this with 98.9% giving it a high preference.

The use of Internet or digital devices to recite the Quran also helps people use other facilities or services that are available within online Quranic/Islamic websites and/or digital apps. Table 8 shows that besides reciting the Quran, reading the Tafseer (Quran explanation) and using search tools are the most used services. Besides the services provided in Table 8, other services include learning Quranic Arabic, reciting Zikr (supplications), listening to YouTube, finding prayer times and Qibla direction, preparing Islamic lectures and sharing Quranic services on social media. The response count is more than the number of participants since this question allows for multiple answers.

TABLE 8
THE MAIN FACILITIES/SERVICES THE PARTICIPANTS USE FROM WEB AND MOBILE APPLICATIONS

Answer Options	Response Percent	Response Count
Quran text & recitation only	27.3%	143
Tafseer	23.1%	121
Use of those apps to help memorize Quran	7.6%	40
Le’rab (Grammar) of the Holy Quran.	5.0%	26
Search Facility	19.5%	102
Phonetic Search	5.7%	30
Statistical Analysis	4.8%	25
Retrieving, Printing contents or resources.	5.9%	31
Other (Please specify) -----	1.1%	6

The survey also shows that 70.5% of participants are pleased with the features of most online/smartphones Quran apps and 73% are pleased with the quality of information provided by most apps. Therefore, there are many Quran apps and websites which the participants favor to use, however, the mostly mentioned are: Ayat (provided by King Saud University <http://quran.ksu.edu.sa/>), iQuran Lite (An IOS app), Quran Android, Quran Explorer <http://www.quranexplorer.com/quran/>, and Tanzil (<http://tanzil.net/>). To conclude this category the participants were asked to provide their comments/suggestions/vision on how new technologies could be used to develop and improve Quran related smart apps and web applications. Finally, the participants provided their comments and recommendations on how new technologies can be used to develop and improve Quranic related smartphone and web apps. The participants provided many comments, suggestions and recommendations. Some of the design features the participants would like to see in Quran apps are: zooming, contextual search, adjustable font size, high sound quality, multi-language, Tajweed learning, contact a scholar, daily random lessons and stories, pop up Word by word meaning while reciting Quran, Daily alarm/reminder, quiz, game-based feature to keep motivation, statistical analysis, advanced research and navigation features. Other features or apps the participants would like to see for or in Quran apps include, interactive and user-friendliness, offline accessibility, free of charge, robustness, ad free, free of charge, encapsulates

different features in one app such as qibla direction, prayer time ... etc., user friendly voice-driven apps, an app that could match reciter's voice with the database then give rating of the quality of the recitation, features for people with special needs. Regarding security and authentication, the participants emphasized that authentication should be ensured as it increases the confidence of the user with secured apps. In addition, the participants provided the following comments and suggestions:

- Develop one global source of authentic digital Quran provider.
- Use cryptographic algorithms to ensure data integrity.
- Develop a security tool that is compatible with most Quran apps to be used to check originality of the Quran.
- Develop a secure cloud-based repository for Quran related resources and make it available as a web service for all mobile and web apps. Any Quran related resources must be first authentic before being used in designing apps.
- One or more Islamic authentication companies must check all available digital apps and issue them tags for authentication or rejection, which will make the users aware of authentic apps.
- Provide in app capability for real time authentication.
- Design a plug in that checks Quran verses online.
- Develop an agent program to keep monitoring the Quran text for correctness. Such text may be any uploads related to Quran with the help of popular social media and other websites that may contain such contents.

B. Software/SmartPhone/Digital Application Developer

The list of questions for this category is provided in Table 9. The results provide the users response to questions related to their personal experience as web and mobile application developers.

TABLE 9
THE "SOFTWARE/SMARTPHONE/DIGITAL APPLICATION DEVELOPER" CATEGORY QUESTIONS

No.	Question	Response Count
1	Are you involved in Digital Application Development for Smart devices or Web-Apps?	64
2	Did you ever develop any online Quran application that involved reading or reciting Quran using a Smart device?	64
3	Do you think the digital Quran which you have used in your application is 100% authentic and is free of any intentional/unintentional errors?	63
4	Do you consider "Content or Quranic Script Integrity/Authenticity" as a Primary Measure while developing a Quran application?	63
5	Do you apply Modern Image Processing or Signal processing techniques to prevent any tampering attempts?	62
6	What type of Signal-Processing or Image-Processing tools do you use?*	53
7	Do you follow any particular information security standards while developing applications?*	60
8	Which Security Standards do you follow? (If the previous answer is YES)*	26
9	Do you consider Secure Coding while developing such applications?	60
10	Did you ever witness any case in which the integrity/security of a Digital Quran application was compromised due to unsecure coding or development?*	61
11	Do you follow Quality Assurance standards while developing such applications?	59
12	Which Quality Standards do you follow? (If answer for the above question is YES)*	27
13	Will you consider to apply Security Standards, Techniques and Measures to prevent any type of forgeries while developing Quran Applications?	60
14	Do you think it is necessary for a Quran Authentication Body to monitor and endorse the digital copies of the Qur'an Worldwide?	61
15	Will you prefer validating your Quranic resources through a Quran Authentication Body?	61
16	What is your vision/comments/recommendations of how new technologies could be used to develop and improve Quranic/Related Smartphone Apps, Web Apps or user interaction further? Please specify:	26

*Question has an "Other" option or "comment box"

The results of the survey show that 73.4% of the participants are involved in Digital Application Development for Smart devices or Web-Apps and 28.1% developed online Quran application that involved reading or reciting Quran using a Smart device. Approximately One third of the participants think the digital Quran which they have used in their applications is 100% authentic and is free of any intentional/unintentional errors, and 7.9% use no verification procedures during apps development. On the other hand, 42.9% are “Not sure” and 17.5% “Never thought about it”. However, 90.5% of the participants consider Content or Quranic Script Integrity/Authenticity as a Primary Measure while developing a Quran application.

Modern image processing is one of the main technologies that can be used in designing applications and only 22.6% of the participants have used these tools to prevent any tampering attempts. The response rate to this question is very low given that fact that 51.6% responded with “No” and 25.5% “Never thought about this issue”. This question is answered by developers who may or may not have designed Quran applications, and therefore the response rate may not be reflective of the actual developers who designed Quran apps before. The image processing tools used by the participants include MATLAB image processing Tool-Box (30.2%), open source tools (41.5%), proprietary tools such as Adobe (9.4%) and in house dedicated tools (18.9%).

In response to the question “Do you follow any particular information security standards while developing applications?” 28.3 of the participants follow IS standards, 56.7% do not follow any standards, and 15% “Never thought about using any standards”. The list of standards used is provided in Table 10. The response to the “Other” option includes. mathematical criteria, HCI standards, Nist, rfc, dedicated in house standards. Following this, the survey asked if the developers use secure coding while developing such applications, 46.7% responded with “Yes”, 23.3% responded with “No” and 30% responded with “I never thought about secure coding.

TABLE 10
THE LIST OF SECURITY STANDARDS USED BY PARTICIPANTS

Answer Options	Response Percent	Response Count
ISO 9001	50.0%	13
ISO 27001	7.7%	2
PCI-DSS	0.0%	0
FFIEC	0.0%	0
Other (please specify)	42.3%	11

The security of the digital Quran should not be compromised when designing Quran applications. Therefore, security and quality assurance standards are very important in designing software. The results show that only 4.9% of the participants have witnessed some digital Quranic apps with modified/forged due to unsecure coding or development and in response to “Do you follow Quality Assurance standards while developing such applications?” 35.6% say “Yes”, 40% say “No” and 23.7 say “I don’t Know”. The quality standards followed by the participants are given in Table 11. The “Other” option include, mathematical criteria, dedicated, and I use the Quranic text that is known to have been validated fromTanzil.org. Also, 86.7% of the participants expressed their interest to consider applying Security

Standards, Techniques and Measures to prevent any type of forgeries while developing Quran Applications. Also, 98.4% think it is necessary for a Quran authentication body to monitor and endorse the digital copies of the Qur'an worldwide and 95.1% prefer validating Quranic resources through a Quran Authentication Body.

TABLE 11
THE LIST OF QUALITY STANDARDS USED BY PARTICIPANTS

Answer Options	Response Percent	Response Count
ISO 9004	51.9%	14
ISO 19011	3.7%	1
AS9100	0.0%	0
MBNQA	3.7%	1
Lean/Six Sigma/TQM	3.7%	1
McCall Quality Factors	0.0%	0
Other (please specify)	37.0%	10

Finally, the participants provided their vision, comments and recommendations of how new technologies could be used to develop and improve Quranic/Related Smartphone Apps, Web Apps or user interaction. Some of the participants comments and recommendations are:

- Develop interactive apps.
- Provide technical support for monitoring and addressing concerns from users.
- Give more attention to information security and quality assurance.
- Provide the Quran text as a service for developers in a way that they can only integrate it without having to modify.
- Watermarking can be used to tackle the security issue.
- Protect Quran applications by using information security algorithms.
- Follow very rigorous information scrutiny and authenticity procedure for the content by performing intrusion testing to make the application secure.
- Follow well known and secure programming methodologies.
- Develop an intelligent Quran content checker.
- Quranic Text and Translations must be verifiable by the user of an app.
- Anti-tampering feature with the originality of Quran should be included in such application to alert the users.
- Regulatory bodies (in collaboration with scholars) should be formed to monitor the development and distribution of Quran app and empowered in order to take action in case of a tampering case is detected.
- Authentic Digital Quranic text should be made available to Quran digital developers. It should be in form that is flexible enough to be used in all platform.
- Using mobile solution networks the authentic Quranic content can be disseminated among the social users of the communities.

C. Graphic Designer

The list of questions for the Graphic Designer category is provided in Table 12. This category received only 4 responses even though 9 participants chose this category as the one they belong to. The results show that 50% of the participants role involve designing graphics and/or user interfaces for Quranic applications. However, none use or apply secure design concepts while designing Quranic applications for sensitive

scripts such as Quran with one response (25%) was “never thought about this issue.” Also, 50% didn’t ever encounter any case in which the security of a digital Quran application was compromised due to unsecure design and 50% response was “never thought about this issue.” As a result of the above responses all participants agree on the following points:

- They will consider and apply design security standards, techniques and measures to prevent any type of forgeries/tampering in the design of the Quran applications.
- It is necessary for a Quran body authentication body to monitor and endorse the digital copies/publications of the Quran worldwide.
- Prefer to validate Quran resources through a Quran authentication body.

Finally, when the participants were asked about providing comments/suggestions/vision as graphics designers only one response was received which suggested that social media can be linked to digital apps to encourage communications between people from all over the world can talk and discuss things about the Quran, also, there should be an option to contact scholars within such applications.

TABLE 12
THE “GRAPHIC DESIGNER” CATEGORY QUESTIONS

No.	Question	Response Count
1	Does your role involve designing Graphics or User Interface for Quranic applications?	4
2	Do you apply Secure Design concepts while designing Quranic application for sensitive Quranic Scripts?	4
3	Which Secure Design Concept do you use? (If Q.2 is Yes)*	3
4	Did you ever encountered any case in which the security of a Digital Quran application was compromised due to unsecure design?*	4
5	Will you consider and apply Design Security Standards, Techniques and Measures to prevent any type of forgeries/tampering in the design of Quran Applications?	3
6	Do you think it is necessary for a Quran Authentication Body to monitor and endorse the digital copies/publications of the Qur’an Worldwide?	4
7	Do you prefer validating your Quranic resources through a Quran Authentication Body?	4
8	What is your vision/comments/recommendations of how new technologies would be used to develop and improve Quranic/Related Smartphone Apps, Web Apps or user interaction further?	1

*Question has an “Other” option or “comment box”

D. Organization Manager/Director/Decision Makers/Owners/CEOs

This list of questions for the Organization Manager/Director/Decision Makers/Owners/CEOs category is provided in Table 13. The results show that approximately 60% of the participants are currently managers, directors, CEOs or decision makers of companies/organizations. Table 14 shows the current managerial role of the participants. The highest response “Mangers” option received 32.1% followed by “Other” option 29.5%. The “other” option included the following responses: deans of colleges, deputy deans, head of department, supervisors, professors, project managers, IT technologist, lecturers, and salespersons. 64.9% of the respondents are involved in Managing, Directing and Decision Making in Application Development. 75% of the companies work with smart devices and digital application development and only 15.3% of the companies develop and design digital Quran applications.

The results show that 65.2% of the digital Quran content is obtained from online free resources, 13% from Islamic bodies, 7.2% from Islamic institutes and 14.5% of the participants are not sure. In addition, 20.6% mentioned that they carry proof of authenticity for the Quran copies they develop without providing

any details, however, 79.4% do not have digital proof or never thought about Quran authentication. In response to the question “Does your company follow any particular security standards while developing these applications?” 34.3% responded “Yes” and 65.7% responded with “No” or “Never thought about using such standard.” Table 15 provides the security standards followed by these companies. The “Other” option included ISO9001 + CMMI for DEV. L2, and SSL.

TABLE 13
THE “ORGANIZATION MANAGER/DIRECTOR/DECISION MAKERS/OWNERS/CEOS” CATEGORY QUESTIONS

No .	Question	Response Count
1	Are you the current Manager/Director/CEO/Decision Maker/Owner of your company/Organization?	76
2	What is your current Role?*	61
3	Are you involved in Managing, Directing and Decision Making in Application Development?	74
4	Does your company deal with Smart Devices/Digital Application Development?	72
5	Does your company develop and design Digital Quran Applications?	72
6	From where do you obtain the Digital Quran content?	69
7	What Proof-of-Authenticity do you have that the Digital Quran Copy is/was authentic?	68
8	Does your company follow any security standards while developing these applications?	70
9	Which Security Standards does your company follows? (If the previous answer is YES)*	56
10	Does your company follow Quality Assurance standards while developing such applications?	67
11	Which Quality Standards does your company follow? (If the previous answer is YES)*	50
12	Will you consider adopting Security Standards and Quality Assurance Standards while developing future digital applications?	66
13	Do you think that it is necessary for a Quran Authentication Body to monitor and endorse the digital copies/publications of the Qur’an Worldwide?	69
14	Will you prefer/recommend that your company will seek validating of Quranic resources through a Quran Authentication Body to obtain a Digital Quran Certificate?	70
15	What is your vision/comments/recommendations of how new technologies would be used to develop and improve Quranic/Related Smartphone Apps, Web Apps or user interaction further?	29

*Question has an “Other” option or “comment box”

TABLE 14
THE CURRENT MANAGERIAL ROLE OF THE PARTICIPANTS

Answer Options	Response Percent	Response Count
Manager	32.1%	25
Director	21.8%	17
CEO	9.0%	7
Decision Maker	6.4%	5
Owner	1.3%	1
Other please state)	29.5%	23

TABLE 15
THE SECURITY STANDARDS FOLLOWED BY THE COMPANIES OF THE PARTICIPANTS

Answer Options	Response Percent	Response Count
ISO 9001	17.9%	10
ISO 27001	3.6%	2
PCI-DSS	0.0%	0
FFIEC	0.0%	0
I don’t know	71.4%	40
Other (please specify)	7.1%	4

In response to the question “Does your company follow Quality Assurance standards while developing such applications?” 41.8% of the participants responded “Yes”, 25.4% responded “No” and 32.8% chose “I do not know”. Table 16 provides the Quality Assurance standards followed by these companies. The “Other” option includes: BRC & ISO 9001, company standards, and WISE (in house developed method).

TABLE 16
THE QUALITY ASSURANCE STANDARDS FOLLOWED BY THE COMPANIES OF THE PARTICIPANTS

Answer Options	Response Percent	Response Count
ISO 9004	14.0%	7
ISO 19011	8.0%	4
AS9100	0.0%	0
MBNQA	0.0%	0
Lean/Six Sigma/TQM	0.0%	0
McCall Quality Factors	0.0%	0
I don't know	68.0%	34
Other (please specify)	10.0%	5

“Will you consider adopting Security Standards and Quality Assurance Standards while developing future digital applications” 79.4% say “Yes”, 2.9% say “No” and 17.6% say “I don’t know”. A response with “I don’t know” implies that such participants are not technically oriented or they do make decisions in the company.

“Do you think that it is necessary for a Quran Authentication Body to monitor and endorse the digital copies/publications of the Qur’an Worldwide?” 88.5 % responded Yes with high preference, 10.1 say “Not sure” and 1.4% say “No”. In response to “Will you prefer/recommend that your company seek validating of Quranic resources through a Quran Authentication Body to obtain a Digital Quran Certificate?” 84.3% responded with “Yes”, 12.9% “Not sure” and 2.9% “No”. From the results, it is very clear that many of the participants in this category are not familiar with security and Quality Assurance standards.

Finally, the participants provided their vision/comments/recommendations of how new technologies would be used to develop and improve Quranic/Related Smartphone Apps, Web Apps or user interaction. The following list are the comments recorded by the participants,

- Multilanguage text search and translations.
- Validation of Quran from original source.
- Interactive voice commands
- Need to validate all Quranic information received from the Internet.
- Use data and text mining techniques
- Use NLP techniques as a supporting method in understanding the Holy Quran
- Provide clear, concise and uniform information on Islamic and Quranic issues.
- Authenticity is a must for Quranic applications.
- Recommend a single Islamic institution to develop and implement steps of policies, strategies, standards, procedures, technologies, tools, applications and implementations on international level.
- Support the development of an Islamic accreditation center which will have the authority to issue accreditation certificates for any Quranic/Islamic apps.
- There is an urgent need for authentic Quranic apps for new Muslims especially in Europe.
- Apply cloud computing and semantic web technologies as they are fertile fields for developing Arabic language and Quran apps, through research based on ontology relations and semantic fields in Arabic.
- Develop a matrix of basic standards for quality control and accreditation methodology for the Holy Quran Information Systems (HQIS) based on the "Quality Benchmarks" and "Accreditation" benchmarks. Develop a set of Key Performance Indicators (KPIs), which include a set of key factors and sub-functions. One of the systematic means of testing is to evaluate the basic stages of the Quran software in all its details and components, in order to improve the software and integrate IT in the service of Quranic sciences so that these techniques are integrated into the (HQIS). So, that to qualify

the product to obtain a Quality Mark and to grant the manufacturer an Accreditation Certification, by a recognized donor, so that this matrix is the basis for anyone who wishes to produce software tools in the science of the Quran and is compatible with science and Sharia, whatever the form, type or structure of the program is.

- To ensure systematic dissemination of contemporary techniques used in the sciences of the Holy Quran in all its branches and elements, it is necessary to establish a comprehensive system of quality unified and acceptable from all specialized centers to evaluate and adopt the quality of the software tools related to the Holy Quran. Thus, combining the efforts between the researchers/workers in the field of IT, specialists in the forensic Quranic studies, experts in the field of international standards and software standards in order to develop a matrix of basic criteria for quality control and accreditation methodology for the software tools related to Quranic science. KPIs are used as one of the systematic means to test, measure and evaluate all the basic phases of the Holy Quran software in all its details and components. Therefore, the performance indicators are a set of key factors and sub-functions, which ensure quality controls and accreditation methodology for the Quran science software tools to improve them and integrate IT in the service of Quranic sciences, so that these techniques are integrated and linked to a vital link. To evaluate the accuracy, excellence and efficiency of these software tools, the quality controls and the accreditation methodology for the companies producing the Quran-related software tools will aid in achieving the required level of quality and meet the established standards regarding the organizational structure, procedures, processes and resources required to implement a comprehensive quality management system.

E. Islamic Institutes/Islamic Research Centers/Religious Body/IT or Other Research Centers

The list of questions for the “Islamic Institutes/Islamic Research Centers/Religious Body/IT or Other Research Centers” category is provided in Table 17.

TABLE 17
THE “ISLAMIC INSTITUTES/ISLAMIC RESEARCH CENTERS/RELIGIOUS BODY...” CATEGORY QUESTIONS

No.	Question	Response Count
1	Are you a representative of your institute/organization/center?	20
2	Does your curriculum cover Quranic studies at a Department or College level?*	17
3	Does your institute have Scientific Research related to Quran?	20
4	Does your institute have Research Centers dedicated for Quranic studies and research?	20
5	Does your institute address Digital Authentication and Certification issues in Quranic research and studies?	20
6	Do you think that the Digital Quran copies/publications available online or on different devices are all 100% authentic?	19
7	Based on the Quranic research and academic activities conducted in your institute, do you think that it is necessary for a Quran Authentication Body to monitor and endorse the digital copies/publications of the Qur'an Worldwide?	20
8	Will you prefer that your institute will become a part of such initiatives to secure and validate the digital Quranic resources by forming a Digital Quran Authentication Body and issue a Digital Quran Certificate that confirms the authenticity of the Quran contents?	20
9	What is your vision/comments/recommendations of how new technologies would be used to develop Quranic/Related Smartphone Apps, Web Apps or user interaction further?	10

*Question has an “Other” option or “comment box”

This category received 20 responses, 85% of the participants are representative of their institutes/organizations however, 15% are representatives of non-Islamic institutes/organizations. 85% of these institutes/organizations provide Quranic studies curriculum at a college or department level and 15% at a high school or grade school level. Table 18 provides the responses to 5 of the questions shown in Table 17. The results show that 65% of the research in these institutes is dedicated to Quran with 55% of these

institutes having special research centers for Quranic studies. However, even though most of these research institutes carry research on such sensitive script the issue of authentication is not given high importance, since only 25% of these institutes address the issue of digital authentication and certification. The 15.8% of the respondents believe that the digital copies of the Quran available online or on different devices are 100% authentic. Therefore, 80% of the participants prefer that their institute will become a part of such initiatives to secure and validate the digital Quranic resources by forming a Digital Quran Authentication Body and issue a Digital Quran Certificate that confirms the authenticity of the Quran contents.

TABLE 18
THE RESPONSE TO SOME OF THE QUESTIONS IN THIS CATEGORY

Question	Yes	No	Not sure
Does your institute have Scientific Research related to Quran?	65%	35%	0%
Does your institute have Research Centers dedicated for Quranic studies and research?	55%	45%	0%
Does your institute address Digital Authentication and Certification issues in Quranic research and studies?	25%	55%	20%
Do you think that the Digital Quran copies/publications available online or on different devices are all 100% authentic?	15.8%	21.1%	63.1%
Will you prefer that your institute will become a part of such initiatives to secure and validate the digital Quranic resources by forming a Digital Quran Authentication Body and issue a Digital Quran Certificate that confirms the authenticity of the Quran contents?	80%	15%	5%

Finally, from the comments of the participants it can be concluded that:

- Apps should provide certification and guidance to users.
- Apps should be designed with age groups in mind.
- Apps should be user-friendly
- Authentication is a must.
- Develop Authentication filters to check all information on web.
- ISO certification.
- Use such applications in teaching the Holy Quran.

F. Holy Quran Publishers

The list of questions for the Holy Quran Publishers is given in Table 19. It shows that the maximum number of responses recorded for the Holy Quran Publishers category is 4. The results show that all participants agreed on the following points:

- Their companies publish and distribute Quran related digital content or printed copies in addition to other non-Quranic publications.
- Publication of digital Quran in different media such as smart devices and online are effective and easy methods for users to read Quran.
- Current digital copies of the Quran are free from any intentional or unintentional errors.

Other results from the responses received include:

- 66.7% think that the digital copies of the Quran are all 100% authentic and the participant's organizations can rely on its authenticity; however, 33.3% are not sure.

- In responding to the question “Will your publishing organization consider publishing digital copies of the Quran?” three responded to this question and each one chose a different response, these are: “Yes”, “No”, and “I do not know.”
- Most of the participants (3 out of 4) are unaware if any authentication body is equipped and trained enough to conduct such authentication and validation of digital Quran/resources/publications on the other hand only 1 respondent agreed on this.
- 75% of the 4 participants agree that it is a very high preference/requirement for digital copies of the Quran to be monitored and endorsed by a Quran authentication body and 25% thinks otherwise.
- 66.7% (3 out of 4 participants) would prefer/recommend that their publishing company will consider applying for validating their Quranic resources through a Quran Authentication Body to obtain a Digital Quran Certificate that confirms the authenticity of the digital Quran contents.

TABLE 19
THE “HOLY QURAN PUBLISHERS” CATEGORY QUESTIONS.

No.	Question	Response Count
1	Does your company deal with Publishing and Distributing the glorious Quran? (Digital or Printed)	3
2	Does your company ONLY deals with Publishing and Distributing the printed and non-digital copies/publications of Quran?	3
3	Is there any Authentication Body that deals in authenticating and validating the Quran before final publishing and further distributing?*	3
4	Does your company consider that Digital Quran available in Smart Devices or Online provides another effective and easy method for users to read Quran?	3
5	Do you think that the current digital copies of the Quran is free from any intentional or unintentional errors?	3
6	Do you think that the digital copies of the Quran are all 100% authentic and your publishing organization can rely on its authenticity?	3
7	Will your publishing organization consider publishing digital copies of the Quran?	3
8	Do you think that any Authentication Body is equipped and trained enough to conduct such authentication and validation of digital Quran resources/publications?	4
9	Do you think that it is necessary for a Quran Authentication Body to monitor and endorse the digital copies of the Qur’an Worldwide?	4
10	Will you prefer/recommend that your publishing company will consider applying for validating their Quranic resources through a Quran Authentication Body to obtain a Digital Quran Certificate that confirms the authenticity of the digital Quran contents?	3
11	What is your vision/comments/recommendations of how new technologies could be used to develop and improve Quranic/Related Smartphone Apps, Web Apps or user interaction further?	1

*Question has an “Other” option or “comment box”

Finally, when asked for their comments, recommendations of how new technologies could be used to develop and improve Quranic related smartphone apps, or web apps only one participant suggested to apply Artificial Intelligence for understanding the Quran.

IV. RECOMMENDATIONS

Based on the comments, suggestions and recommendations provided by the participants in different categories surveyed, the main recommendations can be summarized as follows:

- Establish an international Islamic accreditation center to be the only source for publishing the printed and digital copies of the Quran. Its role will also include training, monitoring, endorsing, and scrutinization.
- Support the development of an Islamic accreditation center which will have the authority to issue accreditation certificates for any Quranic/Islamic center so that they can act as certification bodies under the umbrella of the main Islamic accreditation center.

- The International Islamic accreditation center must set a framework with clear milestones for developing standards the issuance of certifications for centers as well as for the design of web and mobile Quran apps.
- This establishment/organization is also responsible to provide certificates of authorization to establish centers in different countries that follow strict rules in publishing Quran printed copies and issue certifications for apps.
- Develop a matrix of basic standards for quality control and accreditation methodology for the Holy Quran Information Systems (HQIS) based on the "Quality Benchmarks" and "Accreditation" benchmarks.
- Develop a set of Key Performance Indicators (KPIs), which include a set of key factors and sub-functions. One of the systematic means of testing is to evaluate the basic stages of the Quran software in all its details and components, in order to improve the software and integrate IT in the service of Quranic sciences so that these techniques are integrated into the (HQIS). So, that to qualify the product to obtain a Quality Mark and to grant the manufacturer an Accreditation Certification, by a recognized donor, so that this matrix is the basis for anyone who wishes to produce software tools in the science of the Quran and is compatible with science and Sharia, whatever the form, type or structure of the program is.
- Apply cloud computing and semantic web technologies as they are fertile fields for developing Arabic language and Quran apps, through research based on ontology relations and semantic fields in Arabic.
- Develop a secure cloud-based repository for Quran related resources and make it available as a web-service for all mobile and web apps. Any Quran related resources must be first authentic before being used in designing apps.
- Provide in app capability for real time authentication.
- Develop an agent program to keep monitoring the Quran text for correctness. Such text may be any uploads related to Quran with the help of popular social media and other websites that may contain such contents.

In regard to features the participants would like to see in mobile and web apps, the following list provides some of the ideas presented by them:

- Social media can be linked to digital apps to encourage communications between people from all over the world can talk and discuss things about the Quran.
- Option to contact scholars within the app.
- Develop interactive apps.
- Provide in-app technical support for monitoring and addressing concerns from users.
- Develop an intelligent Quran content checker.
- Anti-tampering feature should be included in such applications to alert the users.
- Offline accessibility.
- Provide free of charge apps and ad free apps
- Encapsulates different features in one app such as qibla direction, prayer time ... etc.,
- User friendly voice-driven apps.
- Develop apps targeting people with special needs.

V. CONCLUSION

This study presented the results of a questionnaire survey distributed to 500 participants to investigate the aspects of security and authentication, certification and standardization of Quran related applications. The survey provided the feedback/perspective of six different categories of participants and this is the first of its

kind study/survey carried out in Quran computing to confront the issues of authentication, certification and standardization. The participants pointed out that there are few incidents that confirm the existence of errors in few digital versions of the Holy Qur'an that are deployed via the Internet and/or mobile apps and such intentional or unintentional forgery is not acceptable even if it is a small error such as a removal or addition of one letter or diacritic, thus this will deem the app or digital content distorted and invalid. Therefore, it can be concluded that the main concern is not the development of Quran apps with attractive features or cosmetic user interface. Certainly, the issue is the available and standardization of properly/accurately digitized Quran related content/resources that are well authenticated (by scholars) and widely distributed (by recognized bodies). Then the risks in terms of security aspect would be very much reduced and controlled. If this stage is reached then, it will be very efficient and easy to develop sophisticated and advanced apps that can be easily monitored and certified by recognized bodies. The latter is the quality standard that we need to primarily develop for Quran related apps, rather than any other standard that would check/improve the technical aspect of an app. Finally, security and authentication of digital Quran increases the confidence of the user in using Quran apps.

ACKNOWLEDGEMENT

The authors would like to thank and acknowledge the King Abdulaziz City for Science and Technology (KACST) for their financial support for this research work during the academic year 2015/2016 under research grant number "٤٤٠-٣٤ - ا". Also, the authors would like to thank the IT Research Center for the Holy Quran and Its Sciences (NOOR) at Taibah University for providing the facilities to conduct research.

REFERENCES

- [1] Khan, Muhammad Khurram, and Yasser M. Alginahi. "The holy Quran digitization: Challenges and concerns." *Life Science Journal* 10.2 (2013): 156-164.
- [2] Tayan, O. and Alginahi, Y., "Information and Communication Technologies in the Teaching and Spreading of the Holy Quran", Symposium on the Glorious Quran and Contemporary Technologies, Information Technology, October 13-15, Madinah, Saudi Arabia, 2009.
- [3] Tayan, Omar, and Yasser M. Alginahi. "A review of recent advances on multimedia watermarking security and design implications for digital Quran computing." In *Biometrics and Security Technologies (ISBAST)*, 2014 International Symposium on, pp. 304-309. IEEE, 2014.
- [4] Tayan, Omar. "Concepts and Tools for Protecting Sensitive Data in the IT Industry: A Review of Trends, Challenges and Mechanisms for Data-Protection." *International Journal of Advanced Computer Science and Applications* 8, no. 2 (2017).
- [5] Hakak, S., Kamsin, A., Tayan, O., Idris, M. Y. I., Gani, A., & Zerdoumi, S. (2017). Preserving Content Integrity of Digital Holy Quran: Survey and Open Challenges. 7, *IEEE Access*. DOI: 10.1109/ACCESS.2017.2682109.
- [6] Zakariah, Mohammed, Muhammad Khurram Khan, Omar Tayan, and Khaled Salah. "Digital Quran Computing: Review, Classification, and Trend Analysis." *Arabian Journal for Science and Engineering* (2017): 1-26.
- [7] Sabbah, Thabit, and Ali Selamat. "A framework for Quranic verses authenticity detection in online forum." In *Advances in Information Technology for the Holy Quran and Its Sciences (32519)*, 2013 Taibah University International Conference on, pp. 6-11. IEEE, 2013.
- [8] Alsmadi, Izzat, and Mohammad Zarour. "Online integrity and authentication checking for Quran electronic versions." *Applied Computing and Informatics* 13, no. 1 (2017): 38-46.
- [9] Mostafa, Mostafa GM, and Ibrahim M. Ibrahim. "Securing the digital script of the Holy Quran on the Internet." In *Advances in Information Technology for the Holy Quran and Its Sciences (32519)*, 2013 Taibah University International Conference on, pp. 57-60. IEEE, 2013.
- [10] Hilmi, Mohd Faiz, Mohammed Fadzil Haron, Omar Majid, and Yanti Mustapha. "Authentication of electronic version of the Holy Quran: an information security perspective." In *Advances in Information Technology for the Holy Quran and Its Sciences (32519)*, 2013 Taibah University International Conference on, pp. 61-65. IEEE, 2013.
- [11] AlAhmad, Mohammad A., Imad Alshaikhli, and Bashayer Jumaah. "Protection of the Digital Holy Quran hash digest by using cryptography algorithms." In *Advanced Computer Science Applications and Technologies (ACSAT)*, 2013 International Conference on, pp. 244-249. IEEE, 2013.

- [12] Mohammed S. Khalil, Fajri Kurniawan, Muhammad Khurram Khan, and Yasser M. Alginahi, "Two-Layer Fragile Watermarking Method Secured with Chaotic Map for Authentication of Digital Holy Quran," *The Scientific World Journal*, vol. 2014, Article ID 803983, 29 pages, 2014. doi:10.1155/2014/803983
- [13] Kurniawan, Fajri, Mohammed S. Khalil, Muhammad Khurram Khan, and Yasser M. Alginahi. "DWT+ LSB-based fragile watermarking method for digital Quran images." In *Biometrics and Security Technologies (ISBAST)*, 2014 International Symposium on, pp. 290-297. IEEE, 2014.
- [14] Khan, Muhammad Khurram, Zeeshan Siddiqui, and Omar Tayan. "A secure framework for digital Quran certification." In *Consumer Electronics (ICCE)*, 2017 IEEE International Conference on, pp. 59-60. IEEE, 2017.

Proposed Architecture for Improving Security and Consistency of Data Transactions in Cloud Database using Tree-Based Consistency Approach

J. Antony John Prabu,
Research Scholar and Assistant Professor,
Department of Computer Science
St. Joseph's College, Trichy, Tamilnadu, India- 620 002

Dr.S Britto Ramesh Kumar
Assistant Professor,
Department of Computer Science
St. Joseph's College, Trichy, Tamilnadu, India- 620 002

Abstract— Cloud computing is one of the important resources in IT industries to provide different services for cloud vendors and clients. Security is the main issue in cloud because data are stored and maintained by third party environment. Cloud computing has lot of issues to maintain transactional data in cloud databases. It needs to maintain ACID guarantees to execute the transactional data. In this paper, Proposed Architecture, Cloud controller, Depth 1 Fixed Tree Consistency (DIFTC) method, Cloud Data Locker and data storage in cloud have been discussed. It also portrays the needs of ACID guarantees, major security levels and ensure consistency of data transactions. Finally this paper affords thorough study on proposed architecture for improving security and consistency of data transactions in cloud database.

Keywords--- Cloud DTM, 2-Phase Commit Protocol, Cloud Security, Consistency in cloud DBS, Cloud Controller, DIFTC and Cloud Data Locker.

I. INTRODUCTION

Cloud computing provides everything as a service and it has different deployment models. It gives reliable service for analytical data but not for transactional data management [32]. Most of the cloud services are deployed in the hybrid cloud environment. The cloud vendors are making a service agreement with cloud providers and launch the reliable services for their clients. Instead of owning the software and hardware vendors move to cloud environment. So the cloud vendors need not worry about the software and hardware maintenances. Cloud has some technical issues to handle the data in distributed environment [10]. The cloud researchers are designing and developing new techniques for scalable transactions to avoid workload consistent[47,49], concurrency controls to improve serialization access [4, 8, 22], consistency approaches for efficient transaction processing system[24,25,26], scalable data storage model for better data handling[7,12,14], data security models[18,19,21] for efficient data handling and security algorithms for multilevel security. Therefore providing security is one of the common and major concerns in cloud, so multilevel security is needed to protect each service without loss of data [17, 33, 48]. The next major issue is maintaining strong consistency state in the database level especially for transactional data management services [46]. Hence most of the research works are concentrating to develop the reliable methodology and architectures to

strengthen the security for cloud services and strengthen consistency state in cloud distributed database system [15, 16, 30].

II. REVIEW OF LITERATURE

R.Anandhi, et al proposed a model to improve the consistency state of the data transaction and also analyze the performance factors of different scalability options in cloud databases. This paper insists that reliability of cloud transaction applications depend on the consistency and scalability levels [13]. Chang Yao, et al proposed a concurrency control protocol named DGCC ((Dependency Graph based Concurrency Control). This protocol is used to achieve better execution of transaction and scalability among database systems. This DGCC based OLTP system also integrated with efficient recovery mechanisms [1]. Pornpan Ampaporn, et al explore a performance analysis against two different data consistency models by leading cloud providers. The results show that writes performance was 3 times worse than reads and it also has greater variance of consistency rate. Hence a better consistency approach is needed for data transactions in cloud environment [3]. Aleksey Burdakov, et al proposed a consistency model for NoSQL databases for data transactions. It explores the characteristics of data consistency and analysis performance of different consistency models [6]. Álvaro García-Recuero, et al proposed a consistency model to efficiently replicate the data among long geographic distance in cloud environment. This approach secures the overloading of both network and system side. The architecture builds with three dimensional vector field models to handle different applications in cloud [5]. Jens Kohler, et al proposed a architecture called data cache architecture with implementation of both parallel and lazy fetch strategies. This work explores the performance analysis between the two strategies and discusses to overcome the issues with SeDiCo framework [7]. Thuy D. Nguyen, et al proposed a prototype called MLS column-store following kernelized design pattern. This approach used in cloud-scale data storage system. It explores the guarantees of efficient cloud-scale data storage in distributed system [17]. Sebastijan Stoja, et al proposed a architecture for realtime database in cloud data transactions. This paper explores the other important topics and analyzes the merits and limitations of it [38]. Marco Serafini, et al proposed

a dynamic data placement system for partitioned database management system. This system provides ACID guarantees for data transactions. The analyzed results are shown the server capacity and it used to improve the placement quality [38].

III. NEED OF ACID

Ensure guarantee to maintain ACID properties in data transaction is global rule in IT sector. The service providers design efficient architectures to satisfy this issue [27]. Especially in cloud, data are stored in third part environment maintained by cloud vendors and cloud providers [34]. Hence maintaining ACID in cloud is not trivial [28], it is a complex work to implement efficient data transaction services in cloud. Many cloud providers offer their service for data transactions but they have their own merits and limitations [34,35].

The main idea of transaction is sequences of data read and write.

TABLE I. ACID PROPERTIES

A	A transaction executed completely or It goes to the initial state. [All or Nothing]
C	Maintain Consistent state in database
I	The particular transaction will not affect other transactions
D	Ensure a transaction committed successfully

When a transaction committed successfully in cloud environment, it should satisfy the ACID guarantees without loss of any details. The performance analyses depend on the ACID maintenances [36].

IV. EXISTING CONSISTENCY APPROACHES

The frequently used recent consistency approaches in the distributed cloud environment are Classic approach, Quorum approach and Tree-Based approach. The through study of these three consistency approaches are discussed with its merits and limitations.

a) Classic Approach:

The classic approach maintains consistency state through synchronous replication in distributed environment. In this approach all nodes or database servers have participated in the writing operation [2]. Hence, it has low consistency rate and big execution time for each transaction in cloud environment [9, 23].

b) Quorum Approach:

The Quorum approach is stronger than the classic approach that is frequently used in cloud to replicate the cloud databases. The so-called quorum method of voting is used for replication among the cluster of servers in cloud environment. In this approach, the majority of votes from all participated nodes or servers are confirmed for further execution. It shows more performance and gives high consistency assurance than the classic approach, but it slow down during the database execution because of frequent voting system [9, 23].

c) Tree-Based Approach:

The Tree-Based consistency approach is the leading approach to maintain data consistency in the distributed data bases. It is formulated based on complex tree structure. Hence it provides varies degree of consistency rate depends on the replica server placed in the level of trees[2]. This approach introduces components in the cloud environment for the better execution of transactions [9]. The highest level of replica nodes in the tree provides high consistency assurance but it gives low consistency assurance moves to lowest level of replica nodes on the tree [9, 23].

V. SECURITY AT DIFFERENT LEVELS

Multi level security is essential for distributed database system. Every level of protection is important to avoid data loss for efficient data service. More concerns are needed to provide data service in cloud because data are maintained by third party environment [37, 45]. So proper security models are may strengthen the cloud services [20,31].

Cloud services need security at following levels:

TABLE II. V. SECURITY AT DIFFERENT LEVELS

Level of Security	Description
Server access security	It ensures the access control (Authentication, Authorization, and Auditing) to services in the cloud environment.
Internet access security	Connectivity and Open access manage in the public cloud. Infrastructure Security at the Network Level. Ensuring availability of the Internet facing resources of the public cloud used by the organization.
Database access security	Ensuring Access control for database and Key management for encrypting.
Data privacy security	Ensuring data confidentiality and integrity of the organizations data in transit to and from the public cloud provider.
Program access Security	Ensuring access control security for the programs of the client's applications in the public cloud.

VI. DATA TRANSACTION IN CLOUD

a) ElasTraS: An Elastic Transactional Data Store in the Cloud:

ElasTraS is designed for scalable and elastic data transaction in cloud databases[39]. It add components to achieve the elasticity in data storage during data transaction.

It uses two level hierarchy to maintain transaction guarantee and also make elastic scalability while increasing workload. ElasTraS has overcome the limitations of DDBMS with its database techniques for isolation and concurrency control.

b) *G-Store: A scalable Data Store for Transactional Multi key Access in the Cloud:*

G-Store is a scalable data store with multi key access in the cloud environment [40]. It is designed to achieve scalability, availability and fault-tolerance. The key group abstraction procedure is allowed to select any set of key group in the data store and provide scalable transactions. The atomicity and consistency guarantees are maintained by the single key to group of keys .

c) *Scalable Transactions for Web Applications in the Cloud:*

This approach supports scalable transactions in cloud environment [29] . It has transaction manager and many number of local transaction managers to handle the transactions. It maintains the ACID properties even in the server failures. The local transaction managers replicate the data and periodically checkpoint data snapshots to cloud data storage service [30].

d) *EcStore: Towards Elastic Transactional Cloud Storage with Range Query support:*

The EcStore deployed among the cloud cluster to provide high elasticity with efficient range query to support cloud data transactions [41]. It achieves the features like load balancing, data partitioning, data replication and efficient range query for each transactional access. The distributed storage layer, replication layer and transaction manager layers are supported to handle data in cloud storage system.

e) *Dynamo: Amazon's Highly Available Key-Value Store:*

Dynamo is built for amazon to achieve high availability and scalability among the cloud clusters [42]. It takes a step to satisfy high availability, consistency, performance and cost-effectiveness. In this approach data is partitioned and replicated with consistent hashing and consistency is maintained by object versioning. The quorumlike technique is used to maintain consistency state among replicas during transaction updates with decentralized replica synchronization protocol.

f) *Megastore: Providing Scalable, High Available Storage for Interactive Services:*

Megastore is a highly available storage system for interactive services in cloud [43]. Most of the NoSQL storage system like Google's Bigtable and HBase are fully support scalable but they have loose consistency model to maintain consistency state. It satisfy ACID properties over remote replicas with low latency for interactive applications.

g) *Sinfonia: a new paradigm for building scalable distributed systems:*

Sinfonia provides efficient and consistent access of data for mini data transactions to abstract the problem faced from concurrency and transaction failures [44]. It avoids the message passing protocols to minimize the complexity for the development process. The Sinfonia developers manipulate data centre infrastructure system like file system, lock manager and communication services.

VII. PROPOSED ARCHITECTURE

The proposed architecture mentioned in Fig improves security and consistency for data transactions in cloud environment. The components placed in this architecture are discovered after analysis of related works done by the field.

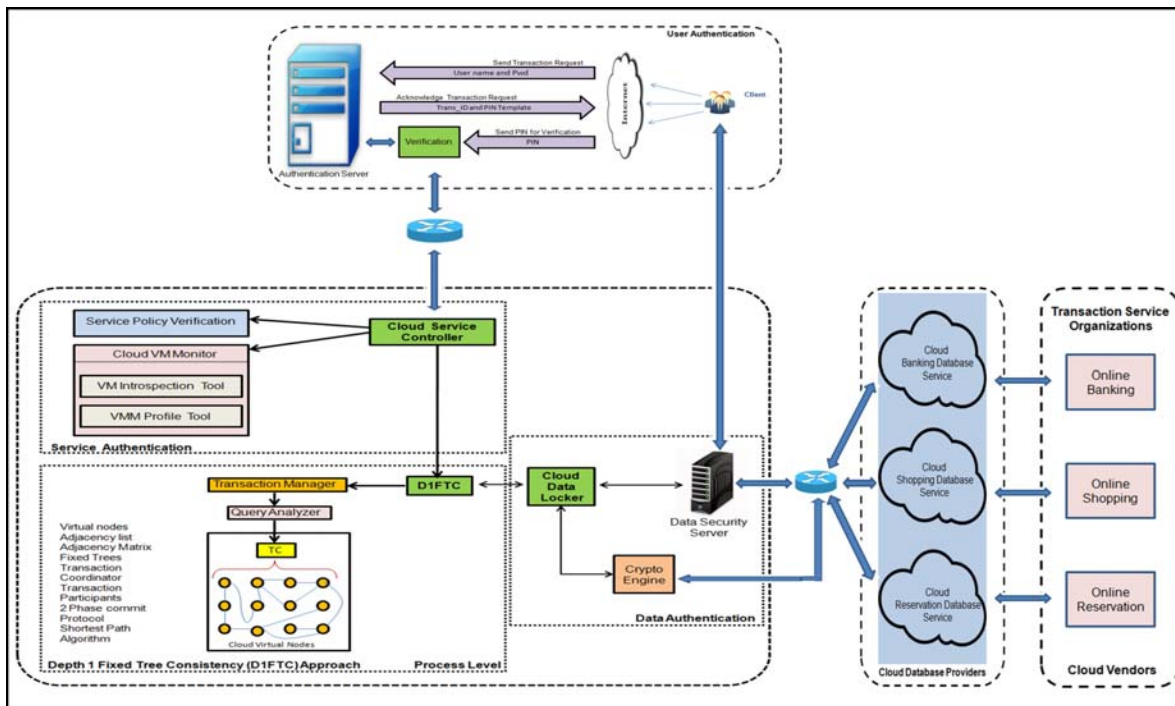


Fig 1. Proposed Architecture

The further execution details are explored thoroughly in this section. main components are mentioned shout.

- a) User Authentication
- b) Cloud Service Controller
- c) Depth 1 Fixed Tree Consistency (D1FTC) Approach
- d) Cloud Data Locker

a) User Authentication:

The client sends the transaction request to authentication with username and password through internet. The authentication server acknowledges the request and sends the transaction ID with PIN template to the client. The client sends the PIN to authentication server for verification. The authentication server verifies whether it is a correct PIN for the specified account or not. After the proper verification process the request may enter to the service level.

b) Cloud Service Controller

The frame work for the cloud service controller makes clear the functionalities of it to enhance the service authentication. After the user authentication the client's requests enter to the cloud service controller. It has two main components named service policy verification and cloud virtual machine monitor.

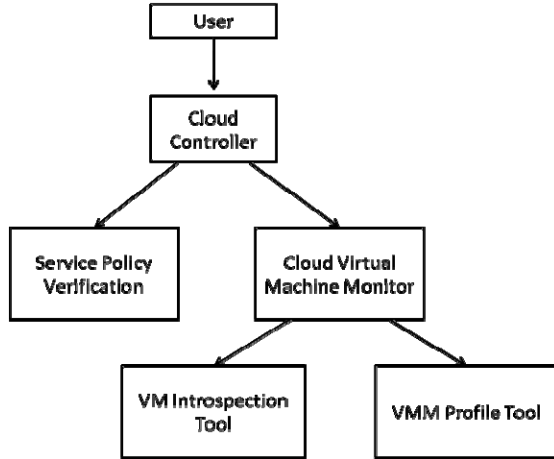


Fig 2. Cloud Service Controller

b.1) Service Policy Verification:

The service policy verification verifies the SLA between cloud providers and cloud vendors. It ensures the Measurement, Condition Evaluation and Management services are not violated. After the service policy verification the request enters in to the virtual machine monitor.

b.2) Cloud Virtual Machine Monitor:

It has two different tools named VM Introspection and VMM Profile tools. Mainly these tools are occupied to verify the status of the each virtual machine in the data transaction management.

b.2.1) VM Introspection Tool:

Number of virtual machines are created and involved in the data transaction execution. The virtual machine introspection tool especially used to evaluate the virtual machines. It frequently verifies the virtual machines and ensures the virtual machine functionalities are good. This

approach is very much useful to maintain the virtual machines for providing reliable service for data transaction clients.

b.2.2) VMM Profile Tool:

Virtual Machine Manager (VMM) contains some of the specifications about the virtual machines. It has virtual machine profiles used to simplify the work to create templates. Templates are used to create virtual machines quickly with proper hardware and operating system settings. The VMM profile has different types of profiles like hardware profile, operating system profile and virtual hard disk profile. These profiles are used to create virtual machine from the created templates.

c) Depth 1 Fixed Tree Consistency (D1FTC) Approach:

The proposed D1FTC method is especially comfort with data transactions in cloud. It efficiently supports 2 phase commit protocol to execute each transaction without affect ACID properties even in the critical situations. The preliminary setup and methodology of the proposed D1FTC method is elaborately explained as follows.

c.1) Functional Procedures:

Preliminary setup:

Step 1: Generate an undirected graph for on hand virtual machines (nodes)

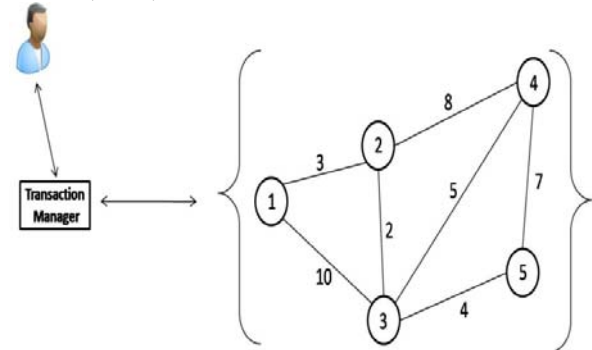


Fig 3. undirected graph for cloud virtual machines with distances

Step 2: Create possible Depth 1 Fixed Trees (D1FTC)

Create adjacency list for the undirected graph and Refer it to find possible depth 1 fixed trees. The number of nodes in undirected graphs is considered as cloud virtual machines.

Adjacency list for the graph is as follows:

Node	Possible Linked Nodes
1	2 3
2	1 3 4
3	1 2 4 5
4	2 3 5
5	3 4

Fig. 4. Adjacency List

The adjacency list specifies the relationship between each node to other nodes in the graph. Hence it very much useful to create possible number of depth 1 fixed trees for the graph as follows:

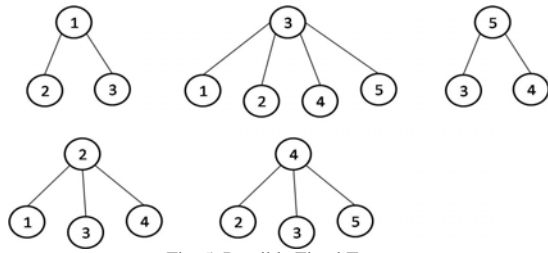


Fig. 5. Possible Fixed Trees

Step 3: Find the distance between each nodes in DIFTC

Create adjacency matrix for the undirected graph and refer it to find the distance between nodes in the tree and count the nodes in each trees. The adjacency matrix for the above mentioned undirected graph is as follows:

	1	2	3	4	5	No. Of Nodes	Distance between Nodes
1	α	3	10	α	α	3	13
2	3	α	2	8	α	4	13
3	10	2	α	5	4	5	21
4	α	8	5	α	7	4	20
5	α	α	4	7	α	3	11

Fig. 6. Adjacency Matrix

Step 4: Find the shortest path from all nodes

The objective is to find the shortest path from each node to all other nodes in the graph. The Dijkstra's Algorithm is suitable way to find the shortest path from one node to other nodes in the graph.

Fig (a), desire to find shortest path from node 1. Edge values in the graph are weights and Node values in the tree are total weights. The transaction manager fixes one nearest node for starting position to find shortest path to connect all nodes. Fig (b) is the shortest path from node 1 and Fig (c) is created as fixed tree for shortest path. It works to update the data among all cloud virtual machines after successful commit of the cloud transaction. So the transaction manager replicated the data after successful commit with shortest path of fixed trees.

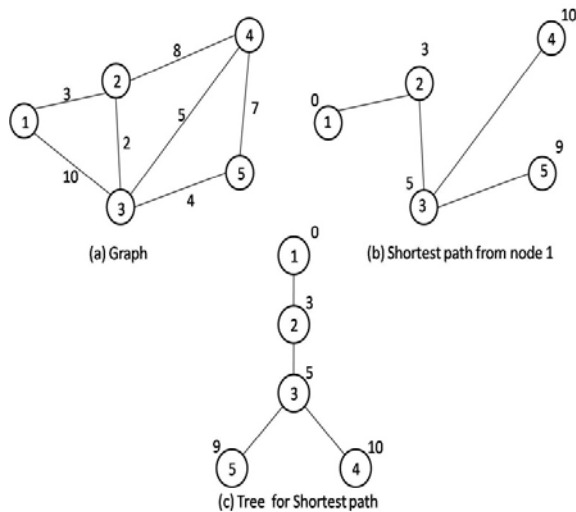


Fig. 7. Find shortest path

DIFTC Method:

The DIFTC method is ready for execution after the proper preliminary works. It handles the virtual machines efficiently to commit each transaction successfully.

The components of DIFTC are as follows:

1. User
2. Transaction Manager
3. Fixed Trees
4. Transaction Coordinator
5. Transaction Participants
6. Two-Phase commit protocol
7. Shortest path Trees

a) User:

The user can submit request to the transaction manager and get response without loss of details.

b) Transaction Manager:

The transaction manager manages all transaction requests from different users and it fix a node for read query and fix a tree for update operations. It analyzes the query and fix the suitable tree with transaction coordinator and participants.

c) Fixed Trees:

The structures of these possible fixed trees are reliable to implement two-phase commit protocol, because it has one to many relationships that is one transaction coordinator and many transaction participants for all update transactions. A participant need not affect the other participants and it communicates only with the coordinator. So the transaction manager can choose any one of the node for transaction coordinator and linked nodes under the transaction coordinator are chosen as transaction participants.

c.1) Adjacency Matrix:

The transaction manager fixes any one of the tree that depends on the required nodes needed to execute a transaction. So the adjacency matrix calculates the number of nodes and the distance between nodes that is to simplify the work of transaction manager.

d) Transaction Coordinator:

Transaction Coordinator is responsible for the given transaction and it maintains all the participants in the selected fixed tree to commit a transaction. The two-phase commit protocol implements in the transaction coordinator.

e) Transaction Participants:

The transaction is divided into small no of process and it sends to transaction participants in the fixed tree. All participants are under supervised by the transaction coordinator.

f) Two-Phase commit protocol:

Two-Phase commit protocol is one of the efficient ways to execute data transactions in the distributed system. It can help successfully and execute the transactions with ACID guaranties in cloud environment. It is also very reliable for the proposed DIFTC method.

g) Shortest path Trees:

In cloud, data replicates from large geographic distance for every transaction and data may be lost during the replication process. The work of the finding shortest path is to connect all

virtual machines. It is to avoid the inconsistency database and minimize the replication time in cloud environment.

c.2) Framework:

The frame work for DIFTC approach is elaborately shown bellow. The step by step process executions are clearly defined and it leads to execute the data transaction and replicate the data in cloud distributed database system.

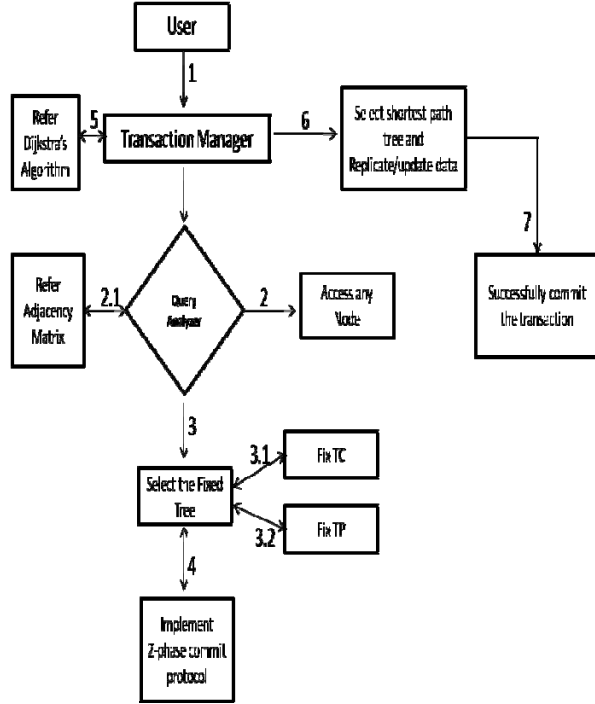


Fig. 8. Frame work for DIFTC

c.3) Pseudocode for DIFTC Approach :

Step 1 : User request sends to transaction manager
 Step 2 : If (Read operation)
 {
 Allocate any node to get updated data.
 }
 Else
 Step 2.1:Refer Adjacency Matrix
 Step 3 : Select the Depth 1 Fixed Tree [According to the transaction query,]
 Step 3.1:Fix the root node as transaction coordinator(TC)
 Step 3.2:Fix the child nodes are transaction participants(TP)
 Step 4 : Implement 2-phase commit protocol in the TC.
 TC divides the transaction and send it to Transaction Participants (TP)
 TC executes the transaction
 Step 5 : Refer Dijkstra's Algorithm
 Step 6 : Select shortest path tree and Replicate/update the data.
 Step 7 : Successfully commit the transaction

d) Cloud Data Locker

In cloud, data are stored in the third party environment and it accessed by remote clients. The cloud providers offer database as a service with their own functionalities. The cloud vendors purchased the Infrastructure, platform, software and database and launch the services for their clients. Hence maintaining security in data storage level is not easy for cloud data transactions. The proposed cloud data locker model has three stage verification.

d.1) Frame work:

The first stage Cloud Data Locker (CDL) model verifies the user to send and get OTP through internet. The second stage it accesses Data Security Server and it verifies the data storage (providers) to send and get the Data Security Number (DSN). The third stage is enabled after the successful verification of both stage one and two. It accesses the Crypto Engine to handle the encrypted data stored in cloud data storage. The crypto engine decrypts the data for execution and encrypts the data after execution to store it in cloud data storage. So the provider does not know which data stored by the clients and the user, vendors are verified with OTP and DSN systems. Hence these three stage verification ensure the high data level security for cloud environment.

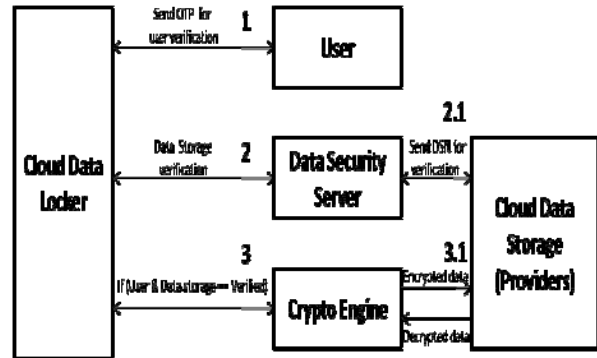


Fig. 9. Frame work for CDL

d.2) Pseudocode for cloud data locker:

Step 1: Send OTP to user
 Step1.2: get OTP from user
 Step1.3: Verify the user
 Step2: Send Data Security Number (DSN) to Cloud Data storage provider
 Step2.1: Data Security Server match the DSN in Data storage provider
 Step2.2: Verify the Cloud Data storage provider
 Step3: If (User & Data storage == Verified)
 Step3.1: Access the Crypto Engine and decrypt the data
 Step3.2: Send data to the transaction process
 Step3.3: Get the committed data
 Step3.4: Access the Crypto Engine and encrypt the data
 Step3.5: Store the data successfully.

VIII. FEATURES OF PROPOSED ARCHITECTURE

a) Security for all levels

The proposed architecture provides multilevel security for every data transactions. User authentication server, cloud controller, cloud data locker are constructed to strengthen the security in user, service and data storage levels.

b) Ensure Service level agreement:

The SLA between cloud providers and vendors are verified continuously to offer reliable service for clients of the vendors. The service policy verification makes sure the SLA is up-to-date.

c) Efficient Consistency Method:

DIFTC is a proposed consistency approach designed for data transaction management in cloud. The adjacency list and adjacency matrix are referred to construct depth 1 fixed trees. It measured the number of nodes and distance between nodes in a fixed tree. After every commit, data are replicated through the shortest path tree referred with Dijkstra's Algorithm.

d) Consistent structure for 2-phase commit protocol:

The depth 1 fixed trees are structured specially to implement two-phase commit protocol. All fixed trees have one transaction coordinator and more than one immediate transaction participants. Hence the DIFTC structure is fully reliable for two-phase commit protocol.

e) Enhanced Data Security:

The proposed Cloud Data Locker model has three stage verification processes to enhance the data storage security in cloud. It verifies the user authentication with OTP system, verify transaction services with data security server and access the data through crypto engine. The data are stored in encrypted form in the cloud. The cloud providers and vendors cannot access or operate the client's data.

f) Guarantee Access Control:

The proposed architecture ensures the access control in all levels. It grants proper implementation of user authentication server, cloud service controller and cloud data locker to manage access control in user, service and data storage level.

g) Ensure ACID guaranties:

The proposed DIFTC approach maintains the *Consistency* state for data transactions in cloud. The two-phase commit protocol is implemented in the fixed trees of DIFTC approach to maintain the *Atomicity* and *Isolation* properties. After the data transaction execution data are updated / replicated through the shortest path tree to maintain the *Durability* of every data transactions. So the proposed architecture ensures the ACID guaranties for data transactions in cloud environment.

h) Efficient data replication:

After every execution of transaction the data should update/replicate with all servers in distributed cloud database. The proposed DIFTC refer the Dijkstra's Algorithm to find the shortest path tree from each node to all other nodes in cloud. The transaction can choose reliable shortest path tree to update / replicate the data efficiently with in the short span of time compare with other approaches.

i) Trusted Transactions:

The cloud service providers (vendors) and end users expect conviction for every transaction. The proposed DIFTC approach has transaction manager to track the transaction from beginning to go to commit state. The transaction coordinators in fixed trees are responsible for allotted transactions. If the system faces any failure, it goes to the initial state without loss of the data.

j) Easy implementation:

The implementation of proposed architecture is simple in the cloud environment. The virtual machines are fixed for transaction process with the reference of Adjacency list, adjacency matrix and Dijkstra's Algorithm in the proposed DIFTC approach. The proposed cloud service controller has the VM introspection and VMM profile tools to monitor the state of virtual machines and maintains them. The proposed cloud data locker has simple three stage methodology to provide high level of data security in the cloud storage. Hence the easy implementation of proposed architecture is reliable for cloud environment.

k) Minimize the response time:

The proposed DIFTC approach designed to implement the two-phase commit protocol to execute transactions faster than other approaches. It allocates the fixed trees depends on the weight of the transactions. It has the shortest path trees to replicate /update the data among cloud virtual machine. So this efficient methodology may minimize the response time for data transactions in cloud.

IX. CONCLUSION AND FEATURE WORK

Cloud provides reliable services to handle analytical data but it faces security and consistency issues when it is offering transactional data management like banking, online reservation and shopping cart, etc. In this paper, the proposed architecture ensures the security in user, service and data storage level with efficient cloud controller and cloud data locker model. The proposed DIFTC approach minimizes the execution time among the virtual machines scatter in the distributed cloud environment. Hence the proposed architecture designed to ensure service level agreement, efficient transaction processing approach, ensure ACID properties, enhanced data storage security for trusted data transactions in cloud environment. The feature work is to implement the proposed architecture in real time cloud environment and do the performance analyze with existing techniques.

X. REFERENCES

- [1] Chang Yao, Divyakant Agrawal, Pengfei Chang, Gang Chen, Beng Chin Ooi, WengFai Wong, Meihui Zhang, "DGCC: A New Dependency Graph based Concurrency Control Protocol for Multicore Database Systems", arXiv:1503.03642v1 [cs.DB] 12 Mar 2015.
- [2] Jasmina Dizdarevic, Zikrija Avdagic, "The aspects of consistency management of highly-distributed transactional database in a hybrid cloud environment for the energy sector", XI International

- Symposium on Telecommunications (BIHTEL) October 24-26, 2016.
- [3] Pornpan Ampaporn and Sethavidh Gertphol, "Performance Measurement of SimpleDB APIs for Different Data Consistency Models", 978-1-4673-7825-3/15/\$31.00 ©2015 IEEE.
- [4] Peter Bailis, Alan Fekete, Michael J. Franklin, Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica, "Feral Concurrency Control: An Empirical Investigation of Modern Application Integrity", Publication rights licensed to ACM, SIGMOD'15, May 31–June 4, 2015.
- [5] Álvaro García-Recuero, Sérgio Esteves, Luís Veiga, "Quality-of-Data for Consistency Levels in Geo-replicated Cloud Data Stores", IEEE International Conference on Cloud Computing Technology and Science, 2013.
- [6] Aleksey Burdakov, Uriy Grigorev, Andrey Ploutenko, Eugene Tsviashchenko, "Estimation Models for NoSQL Database Consistency Characteristics", 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, 2016.
- [7] Jens Kohler and Thomas Specht, "A Performance Comparison Between Parallel and Lazy Fetching in Vertically Distributed Cloud Databases", 978-1-4673-8149-9/15/\$31.00 ©2015 IEEE.
- [8] Peter Bailis, Alan Fekete, Michael J. Franklin, Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica, "Coordination Avoidance in Database Systems", Proceedings of the VLDB Endowment, Vol. 8, No. 3, Copyright 2014 VLDB.
- [9] Md. Ashfakul Islam, Susan V. Vrbsky and Mohammad A. Hoque, "Performance Analysis of a Tree-Based Consistency Approach for Cloud Databases", International Conference on Computing, Networking and Communications, Cloud Computing and Networking Symposium, 2012 IEEE.
- [10] Stefan Schulte, Christian Janiesch, Srikumar Venugopal, Ingo Weber, Philipp Hoenisch, "Elastic Business Process Management: State of the Art and Open Challenges for BPM in the Cloud", arXiv:1409.5715v2 [cs.DB] 22 Sep 2014.
- [11] Palivela Hemant, Nitin P. Chawande, Avinash Sonule, Hemant Wani, "DEVELOPMENT OF SERVERS IN CLOUD COMPUTING TO SOLVE ISSUES RELATED TO SECURITY AND BACKUP", Proceedings of IEEE CCIS2011.
- [12] Divy Agrawal, Amr El Abbadi, "A Taxonomy of Partitioned Replicated Cloud-based Database Systems", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 38(1), March 2015.
- [13] R. ANANDHI, K. CHITRA, "A Challenge in Improving the Consistency of Transactions in Cloud Databases - Scalability", International Journal of Computer Applications (0975 – 8887) Volume 52– No.2, August 2012.
- [14] Joarder Mohammad Mustafa Kamal, Manzur Murshed, Rajkumar Buyya, "Workload-Aware Incremental Repartitioning of Shared-Nothing Distributed Databases for Scalable Cloud Applications", IEEE/ACM 7th International Conference on Utility and Cloud Computing, 2014.
- [15] Ricardo Jimenez-Peris, Marta Patiño-Martinez, Bettina Kemme, "CumuloNimbo: A Cloud Scalable Multi-tier SQL Database", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2015.
- [16] Mohd Muntjir, Mohd Asadullah, Shabbir Hassan, Altaf Hussain, "Cloud Database Systems: A Model Conversion in Databases", 2nd International Conference on Information Technology and Electronic Commerce ICITEC 2014.
- [17] Thuy D. Nguyen, Mark Gondree, Jean Khosalim, and Cynthia Irvine, "Re-thinking Kernelized MLS Database Architectures in the Context of Cloud-Scale Data Stores", Springer International Publishing Switzerland 2015.
- [18] Tahmineh Sanamrad, Lucas Braun, Donald Kossmann and Ramarathnam Venkatesan, "Randomly Partitioned Encryption for Cloud Databases", IFIP International Federation for Information Processing 2014.
- [19] Vasily Sidorov, Wee Keong Ng, "Transparent Data Encryption for Data-in-Use and Data-at-Rest in a Cloud-Based Database-as-a-Service Solution", IEEE World Congress on Services, 2015.
- [20] M. Thamizhselvan, R. Raghuraman, S. Gershon Manoj, P. Victor Paul, "DATA SECURITY MODEL FOR CLOUD COMPUTING USING V – GRT METHODOLOGY", IEEE Sponsored 9th International Conference on Intelligent Systems and Control (ISCO), 2015.
- [21] Chuan Yao, Li Xu, and Xinyi Huang, "A Secure Cloud Storage System from Threshold Encryption" 5th International Conference on Intelligent Networking and Collaborative Systems, 2013.
- [22] Philip A. Bernstein, Sudipto Das, Bailu Ding, Markus Pilman, "Optimizing Optimistic Concurrency Control for Tree-Structured, Log-Structured Databases", SIGMOD'15, May 31–June 4, 2015.
- [23] Md. Ashfakul Islam and Susan V. Vrbsky, "Tree-Based Consistency Approach for Cloud Databases", 2nd IEEE International Conference on Cloud Computing Technology and Science, 2010.
- [24] B. Jeevarani, Dr. K. Chitra, "Improved Consistency Model in Cloud Computing Databases", 2014 IEEE International Conference on Computational Intelligence and Computing Research, 978-1-4799-3975-6/14/\$31.00 ©2014 IEEE.
- [25] Min Shen, Ajay D. Kshemkalyani, and Ta-yuan Hsu, "Causal Consistency for Geo-Replicated Cloud Storage under Partial Replication", IEEE International Parallel and Distributed Processing Symposium Workshops, IEEE, 2015.
- [26] Wenbo Zhu, Murray Woodside, "Tunable Performance & Consistency Tradeoffs for Geographically Replicated Cloud Services (COLOR)", IEEE 2nd International Conference on Cyber Security and Cloud Computing, 2015.
- [27] Ayman E. Lotfy, Ahmed I. Saleh, Haitham A. El-Ghareeb, Hesham A. Ali, "A middle layer solution to support ACID properties for NoSQL databases", Journal of King Saud University – Computer and Information Sciences, 2016.
- [28] Chao Xie, Chunzhi Su, Cody Little, Lorenzo Alvisi, Manos Kapritsos and Yang Wang, "High-Performance ACID via Modular Concurrency Control", SOS'15, October 4–7, 2015, Monterey, CA, 2015.
- [29] Zhou Wei, Guillaume Pierre, Chi-Hung Chi, "CloudTPS: Scalable Transactions for Web Applications in the Cloud", IEEE TRANSACTIONS ON SERVICES COMPUTING, SPECIAL ISSUE ON CLOUD COMPUTING, 2011.
- [30] Ahmad Waqas, Abdul Waheed Mahessar, Nadeem Mahmood, Zeeshan Bhatti, Mostafa Karbasi, Asadullah Shah, "TRANSACTION MANAGEMENT TECHNIQUES AND PRACTICES IN CURRENT CLOUD COMPUTING ENVIRONMENTS: A SURVEY", International Journal of Database Management Systems (IJDBMS) Vol.7, No.1, February 2015.
- [31] Marian K. Iskander, Tucker Trainor, Dave W. Wilkinson, Adam J. Lee, and Panos K. Chrysanthis, "Balancing Performance, Accuracy, and Precision for Secure Cloud Transactions", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 2, FEBRUARY 2014.

- [32] Govind Sing, Manmohan and Garima Tiwari, “ Cloud Computing: A New Era of IT Opportunity and Challenges”, International Journal of Engineering and Management Research, Vol.-3, Issue-4, August 2013.
- [33] Palivela Hemant, Nitin.P.Chawande, Avinash Sonule and Hemant Wani, “DEVELOPMENT OF SERVERS IN CLOUD COMPUTING TO SOLVE ISSUES RELATED TO SECURITY AND BACKUP” Proceedings of IEEE CCIS2011, 978-1-61284-204-2/11/\$26.00 ©2011 IEEE.
- [34] Daniel J. Abadi, “Data Management in the Cloud: Limitations and Opportunities”, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2009.
- [35] Ms. Shalini Ramanathan, Dr.Savita Goel, Mr.Subramanian Alagumalai, “Comparison of Cloud Databases: Amazon’s Simple DB and Google’s Big table” International Conference on Recent Trends in Information Systems, 2011.
- [36] Francis Gropengießer, Kai-Uwe Sattler, “Transactions a la carte – Implementation and Performance Evaluation of Transactional Support on top of Amazon S3”, IEEE International Parallel & Distributed Processing Symposium, 2011.
- [37] Balachandra Reddy Kandukuri, Ramakrishna Paturi V, Dr. Atanu Rakshit, “Cloud Security Issues”, IEEE International Conference on Services Computing, 2009.
- [38] Sebastijan Stoja, Bojan Jelai, Sran Vukmirovi, Darko Capko, Nikola Dalcekovic, “Architecture of Real-Time Database in Cloud Environment for Distributed Systems”, Second International Conference on Artificial Intelligence, Modelling and Simulation, 2014.
- [39] S. Das, A. El Abbadi, C. Science, and U. C. S. Barbara, “ElasTraS_: An Elastic Transactional Data Store in the Cloud” ,USENIX HotCloud 2, 2009.
- [40] S. Das and A. El Abbadi, “G-Store_: A Scalable Data Store for Transactional Multi key Access in the Cloud”, SoCC’2010, Indianapolis, Indiana, USA, 2010.
- [41] H. T. Vo, C. Chen, and B. C. Ooi, “Towards elastic transactional cloud storage with range query support”, Proc. VLDB Endow., vol. 3, no. 1–2, pp. 506–514, Sep. 2010.
- [42] G. Decandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels, “Dynamo_: Amazon ’ s Highly Available Key-value Store”, ACM SIGOPS Oper.Syst. Rev., vol. 41, no. 6, 2007.
- [43] J. Baker, C. Bond, J. C. Corbett, J. J. Furman, A. Khorlin, J. Larson, L. Jean-michel, Y. Li, A. Lloyd, and V. Yushprakh, “Megastore_: Providing Scalable , Highly Available Storage for Interactive Services”, 5th Biennial Conference on Innovative Data Systems Research, 2011.
- [44] M. K. Aguilera, A. Merchant, A. Veitch, and C. Karamanolis, “Sinfonia_: A New Paradigm for Building Scalable Distributed Systems”, SOSP 2007, Stevenson, Washington, USA, 2007.
- [45] Quang Hieu Vu, Maurizio Colombo, Rasool Asal, Ali Sajjad, Fadi Ali El-Moussa and Theo Dimitrakos, “Secure Cloud Storage: A framework for Data Protection as a Service in the multi-cloud environment”, IEEE, 2015.
- [46] Rohan G. Tiwari, Shamkant B. Navathe, “TOWARDS TRANSACTIONAL DATA MANAGEMENT OVER THE CLOUD”, Second International Symposium on Data, Privacy, and E-Commerce, 2010.
- [47] Hatem A. Mahmoud, Vaibhav Arora, Faisal Nawab, Divyakant Agrawal, Amr El Abbadi, “MaaT: Effective and scalable coordination of distributed transactions in the cloud”, 40th International Conference on Very Large Data Bases, September 1st 5th 2014.
- [48] Adam Bates, Ben Mood, Masoud Valafar, and Kevin Butler, “Towards Secure Provenance-Based Access Control in Cloud Environments”, CODASPY’13, February 18–20, 2013, San Antonio, Texas, USA. Copyright 2013 ACM.
- [49] Marco Serafini, Essam Mansour, Ashraf Aboulnaga, “Accordion: Elastic Scalability for Database Systems Supporting Distributed Transactions”, Proceedings of the VLDB Endowment, Vol. 7, No. 12 Copyright 2014.

BIOGRAPHIES



Prof. J. Antony John Prabu is working as an Assistant Professor and pursuing doctor of philosophy in Department of Computer Science, St. Joseph’s College,(Autonomous), Tiruchirappalli, Tamil Nadu,India. He received his M.Phil degree from Jamal Mohamed College, Tiruchirappalli. He received his MCA degree from St. Joseph’s College, Tiruchirappalli. His areas of interest are Cloud Data Transactions and Distributed Technologies.



Dr. S. Britto Ramesh Kumar is working as an Assistant Professor in the Department of Computer Science, St. Joseph’s College (Autonomous), Tiruchirappalli, Tamil Nadu, India. He has published many research articles in the National/International conferences and journals. His research interests are Cloud Computing, Data Mining, Web Mining, and Mobile Networks.

Color Histogram with Curvelet and CEDD for Content-Based Image Retrieval

Heba A. Elnemr, *Department of Computers and Systems, Electronics Research Institute*

Abstract- Content-Based Image Retrieval (CBIR) is one of the most vigorous research areas in the field of pattern recognition and computer vision over the past few years. The accessibility and progressive development of visual and multimedia data, as well as the evolution of the internet, emphasize the necessity to develop retrieval systems that are capable of dealing with a large collection of databases. Many visual features have been explored, and it is virtually observed that implementing one kind of features is not efficient in retrieving different types of images. Therefore, in this paper, the author proposes an efficient image retrieval technique that joins color and texture features. The curvelet descriptors that are obtained by using wrapping based discrete curvelet transform are used as texture features. While color features are extracted using quantized RGB color histogram (QCH). Besides, color edge directivity descriptor (CEDD), which joins color and texture features in one histogram is obtained. A multiclass SVM is applied to classify the query images. Four datasets (ALOI, COIL-100, Wang, and Corel-1000) are used to test and assess the proposed system. Improved retrieval results are obtained over CBIR systems based on curvelet descriptors and CEDD individually and jointly. Furthermore, comprehensive experiments have been performed to select the number of histogram bins that achieves an effective and efficient image retrieval. The obtained average precision for the ALOI, COIL-100, Wang and Corel-1000 datasets are 0.996, 998, 0.898 and 0.964, respectively. Also, comparisons with several state-of-the-arts demonstrate the effectiveness of the proposed system in refining the retrieval performance.

I. INTRODUCTION

Content-based image retrieval (CBIR) is a technique that automatically searches for visually similar images from large scale image databases according to users' requirements. The image retrieval systems based on visual image content have become the center of attention of researchers for more than a decade. The CBIR technique is carried out through two steps; feature extraction and matching policy. In the first step, which considered the most challenging step in CBIR, effective features of each sample image are analyzed and extracted. Most existing general purpose CBIR techniques implement low-level features, such as color, texture, and shape. The set of extracted features is used to build the image signature. While in the second step, the image signatures obtained from the images in the database are compared with that extracted from a query image by a pre-instituted similarity measurement procedure, so that top relevant images in the image database can be returned as the retrieved images [1].

Color content information is one of the most extensively and popularly implemented visual features in CBIR systems. Color feature is comparatively the simplest and most straightforward visual feature for image retrieval. It is also capable of separating images from each other, relatively robust to background complexity and invariant to image size and orientation [2, 3].

The texture is also one of the most commonly utilized low-level visual features in CBIR. Texture features provide spatial and relational information on the intensity distribution over the image [1, 3].

The shape of objects is frequently used as an effective feature for image retrieval because human visual perception can recognize a scene based only on objects shape. Mainly, shape features contain semantic information about an object, thus, the shape description and representation is a very difficult task [3, 4].

During the past years, various techniques have been developed for extracting effective and efficient features. Generally, many developed CBIR techniques are based on extracting a single type of features [4]. However, acceptable retrieval results are hard to be obtained using a single feature type because an image normally

comprises various visual characteristics. Consequently, it is mandatory to combine various feature types in such a way as to enhance and emphasize the quality and efficiency of the extracted features so as to obtain an acceptable retrieval performance [5].

This work aims mainly to construct an efficacious CBIR system that is capable of handling large datasets expeditiously. The proposed CBIR system integrates color and texture features through quantized RGB color histogram (QCH), color edge directivity descriptor (CEDD) and curvelet transform.

RGB color model is the most commonly used color model in CBIR, also a color histogram is the simplest and most widespread color feature utilized for image retrieval. QCH denotes the procedure of reducing the number of histogram bins by gathering similar colors into the same bin. Thus, the QCH has a relatively low computational cost, as well as it is invariant to rotation, translation, and scale. However, QCH does not take into consideration spatial distribution or description of color information, this is deemed to be the main weakness of this method. Also, different images could have similar histograms besides minor variation in color due to variations in luminance could produce a considerable change in histogram [6]. Therefore, the author combines QCH with CEDD and curvelet transform descriptors so as to attain a good performance.

CEDD is a powerful low-level feature that joins color and texture information in a single histogram as well as it has comparatively low computational cost, therefore, it is appropriate to be used in huge image databases [7].

On the other hand, curvelet has been widely adopted for image denoising, character recognition, image segmentation, texture analysis as well as image retrieval and it has shown an encouraging performance [3, 8-12]. As curvelet captures more directional features, besides it grabs more accurate texture and directional information and it outperforms wavelet and Gabor filter [3, 12], the author has implemented curvelet descriptors for her proposed retrieval system. Multiclass SVM model is used to accomplish the classification task.

Finally, the author compared the proposed system with others based on CEDD alone, curvelet alone, and CEDD integrated with curvelet. Furthermore, several investigations have been made to choose the appropriate number of histogram bins to attain an efficient and effective retrieval performance. These features prove to be complementary to each other with promising rendering.

The primary contributions of this work can be summarized as follows. (i) The suggested method extracts color and texture information using CEDD and combines it with texture and directional information that are extracted using curvelet transform descriptors. (ii) The proposed CBIR performance is improved using QCH and various experiments are conducted to determine the best number of bins to achieve the best performance. (iii) The proposed CBIR system performance is examined on several forms of large databases including natural, real world and well defined object images.

The paper is organized as follows: Section 2 presents the related work. Section 3 describes the methodology utilized to build the proposed CBIR system while section 4 portrays the depiction of the empirical setup embracing the datasets and the experimental results. Conclusion and future work are discussed in section 5.

II. RELATED WORK

Latterly, researchers have proposed many approaches for CBIR using different features such as color, shape, and texture. In this section, the author will discuss several recent literatures that covers some key aspects of CBIR technique.

The work of [12] applied discrete curvelet transform on the Brodatz texture images dataset, then low order statistics is computed from the transformed images. Euclidean distance carried out the similarity measurement in

the suggested CBIR scheme. The results manifest that the proposed curvelet texture feature descriptor performs better than that of Gabor filters in both retrieval efficiency and accuracy.

In [13] a retrieval system that uses local feature descriptors; SIFT and SURF, to generate image signatures invariant to scale and rotation is proposed. Then, BOVW model is created by clustering the local descriptors using K-means technique to build the vocabulary of the K clusters. Finally, the retrieval is accomplished using the SVM classifier model.

Ali et al.[14] designed an image representation scheme based on the histograms of triangles. The proposed method aimed to add spatial information to the inverted index of Bag-of-Features representation. Histograms of triangles are carried out through two levels that are evaluated separately. In the first level, the image is divided into two triangles, while in the second level the image is divided into four triangles. Three different classifiers, Radial Basis Function Neural Networks (RBF-NN), SVM and Deep Belief Networks (DBN), are applied and the overall system performance is evaluated.

Authors of [15] presented an experimental study to investigate and analyze the effect of joining four sampling strategies (SIFT, SURF, Random patch generator and Gauss Random patch generator), with four global feature descriptors (MPEG-7 (Color Layout Descriptor (CLD), Edge Histogram Descriptor (EHD) and Scalable Color Descriptor (SCD)) as well as MPEG-7-like (CEDD), in a Bag-Of-Visual-Word (BOVW) structure. The conducted results demonstrate that the retrieval performance of the proposed descriptors outperforms their performance in their original global form. Besides, they perform better than ordinary SIFT and SURF-based approaches and perform comparably or better against some recent methods.

Malik and Baharudin [16] suggested a CBIR technique that is based on extracting quantized histogram statistical texture features in the compressed domain. The grayscale image is divided into non-overlapping blocks. Next, each block is transformed into a DCT block in the frequency domain. The similarity measurement is achieved using seven distance metrics. The experimental results demonstrate that the Euclidean distance has better performance in both computation and retrieval efficiency.

In the work of [17], a CBIR system with texture and color features succeeded by ant colony optimization feature selection technique is proposed. Wavelet transformation of the sample images is computed and the low-frequency components are used as texture features. Dominant color descriptor, color statistic features, and color histogram features are extracted, in both RGB and HSV domains, as color features. For each feature type, a suitable similarity measure is presented. Ant colony optimization technique is implemented to select the most significant features, among the entire extracted features.

Walia and Pal [18] submitted an image retrieval framework based on combining low-level features. The Color Difference Histogram (CDH) is implemented to extract color and texture features, while Angular Radial Transform (ART) is used to extract shape features. The CDH algorithm is modified in order to improve the overall system performance.

In the work of [19], a CBIR system is designed by combining SURF descriptors with color moments. The similarity strategy is carried out by the KD-tree with the Best Bin First (BBF) search algorithm. Voting Scheme algorithm is finally used to classify and retrieve the matched images from the dataset.

Mukherjee et al. [20] proposed a CBIR system relied on assigning a model of visual words to represent an image patch. Each image patch is represented by a vector that denotes the affinities of the patch for the most significant visual words. To improve the retrieval performance a dissimilarity measure among the pair of image

patches is introduced. The dissimilarity measure is made up of two terms: The first one depicts the variation in affinities of the patches that belong to a common set of significant visual words, while the second term penalizes the measure according to the number of visual words that affects only one of the two patches.

III. METHODOLOGY

This paper demonstrates a new CBIR system that relies on extracting texture features as well as color features. The system uses curvelet transform to obtain the spectral domain coefficients that are utilized to compute the texture descriptor of that image while color features are extracted using QCH. Furthermore, CEDD is employed to obtain both color and texture information in a sole histogram. Fig. 1 displays the block diagram of the proposed CBIR system.

A. Quantized RGB Color Histogram

The color histogram is a good portrayal method for describing the color content of an image, it can be obtained by counting the number of occurrences of each color in an image. Pixels in an image are described by three components (typically but not necessarily) in a certain color space, consequently, each pixel is represented as a tuple of three numbers. The RGB color space is the most popular color space used for computer graphics. RGB color histogram investigates each of the RGB-channels separately, this leads to a huge length of the histogram vector ($= 256 * 3$ for 8-bit RGB image). Thus, color quantization has to be applied in order to produce 3D-color histogram, which is suitable for building efficient indexes for large image databases as well as has an acceptable computational cost. In color quantization procedure, the number of colors used to represent an image is reduced, and each color component is quantized into a number of 'bins'. Since color components (R, G, and B) are equally important, each component is quantized into the same number of bins. In this work, the QCH is implemented and a various number of bins are tested in order to find the best quantization level.

B. CEDD histogram

CEDD is a low-level feature that incorporates color and texture information in a single histogram. One of the greatest significant characteristic of CEDD, the low computational power required for its extraction in comparison with that needed for most MPEG-7 descriptors [7, 21]. CEDD is linked with a texture unit to extract textural information besides a color unit to extract color information.

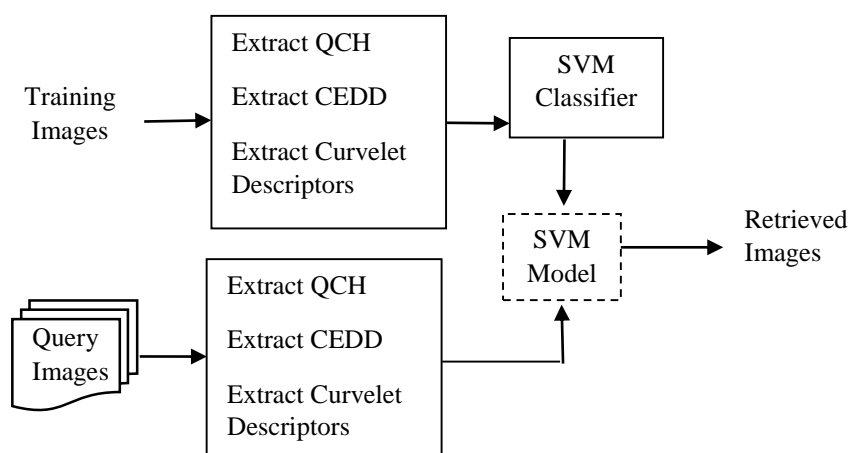


Figure 1. A block diagram of the proposed CBIR system.

In order to obtain the CEDD histogram, each image is divided into 1600 image blocks, and then each block is processed through the following algorithms. In the texture unit, the image block is parted into 4 regions, sub-blocks, the mean value of the luminosity of the pixels enclosed within each sub-block is considered to be the value of this sub-block. The luminosity values are computed within the YIQ color space. Afterward, each block is filtered utilizing 5 digital filters that were suggested in the MPEG-7 Edge Histogram Descriptor [7, 21].

On the other hand, in the color unit, a set of fuzzy rules are implemented to obtain the color information. Each image block is transferred into HSV color space. Next, Fuzzy-Linking histogram technique [22] is applied. The fuzzy system creates a 24-bins histogram.

The CEDD structure incorporates six texture regions and each of which holds 24 color regions. Accordingly, the final histogram consists of $6 \times 24 = 144$ bins, conforming to the total 144 color regions within the six texture regions. Finally, the histogram is normalized and quantized so as each bin is represented by 3 bits. Hence, its total size is limited to be $144 \times 3 = 432$ bits or 54 bytes per image. This small size of the descriptor is considered to be another advantage of CEDD.

C. Curvelet Transform

Curvelet transform is a multi-scale transform designed to represent images at various scales and various angles, this transform is established by Donoho and Duncan in 1999 [23]. In this transform, the input image is initially decomposed into a set of subbands and each subband is partitioned into blocks that are analyzed by ridgelet transform. The ridgelet transform is realized using the Radon transform and the 1-D wavelet transform [24]. To avoid blocking effect, the spatial partitioning process involves overlapping of windows that lead to a large amount of redundancy. As well, it is very time-consuming, which makes it inappropriate for texture feature analysis in a large database. Therefore, Candes et al. [25] developed fast discrete curvelet transform that is based on the wrapping of Fourier samples. This transform is simpler, faster, less redundant, and has less computational complexity since it applies fast Fourier transform instead of the complex ridgelet transform.

For a 2-D input image of size $M \times N$, the curvelet transform based on the wrapping of Fourier samples generates a set of curvelet coefficients indexed by a scale j , an orientation l , and two spatial location parameters (k_1, k_2) . These coefficients are defined as follows [25]

$$C_{j,l}(k_1, k_2) = \sum_{\substack{0 \leq m < M \\ 0 \leq n < N}} f(m, n) \varphi_{j,l,k_1,k_2}^D(m, n) \quad (1)$$

where $f(m, n)$ is the Cartesian array of the input image and $\varphi_{j,l,k_1,k_2}^D(m, n)$ is a digital Curvelet waveform. These coefficients are then used to form the curvelet texture descriptors by implementing statistical operations.

Curvelet Texture Features Extraction

After the curvelet coefficients in each sub-bands are created and stored, curvelet statistical features, i.e. mean and standard deviation, of the coefficients corresponding to each sub-bands, are computed. These features proved to be capable of describing curvelet sub-bands [12, 26]. In general, these mean and standard deviation are then used as texture descriptors of the image. Hence, for each curvelet sub-band, the author obtains two texture features. If n curvelets sub-bands are applied for the transform, $2n$ dimensional texture feature vector is obtained for each image.

TABLE 1
CURVELET TRANSFORM (4 LEVEL DECOMPOSITION)

Scale	1	2	3	4
No. sub-band	1	16	32	1
Sub-bands considered for feature calculation	1	8	16	1

In this work, 4 level curvelet transform is implemented to decompose the input images. Based on this analysis, 50 (=1+16+32+1) curvelet coefficients are computed. However, [12] presents that curvelet at an angle θ generates the same coefficients as curvelet at an angle $(\theta + \pi)$. Thus, due to this symmetry, half of the sub-bands at scale 2 and 3 may be rejected. Accordingly, 26 (=1+8+16+1) sub-bands are maintained, producing a 52 dimension feature vector for each image in the database. Table 1 illustrates the sub-bands distribution at each transform level.

The mean of a sub-band at scale s and orientation θ is stated as:

$$\mu_{s,\theta} = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N C_{s,\theta}(i, j) \quad (2)$$

while the standard deviation of a sub-band at scale s and orientation θ is expressed as:

$$\sigma_{s,\theta} = \sqrt{\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (C_{s,\theta}(i, j) - \mu_{s,\theta})^2} \quad (3)$$

where $C_{s,\theta}(i, j)$ represents the curvelet coefficient at scale s , orientation θ and location (i, j) .

IV. RESULTS AND ANALYSIS

A. Image Datasets

The proposed system was assessed using four different standard datasets; Amsterdam Library of Object Images (ALOI) dataset [27], Columbia object image library (COIL-100) dataset [28] and two subsets of the Corel image database [29].

ALOI image dataset provides one-thousand small objects recorded under various imaging conditions (viewing angle, illumination angle, and illumination color). More than a hundred images of each object were registered. In this work, 102 objects are randomly selected with 72 different viewpoints each. COIL-100 is a well-known standard color image database that includes 100 objects with 72 views acquired by rotating each object around the vertical axis. Two subsets of Corel database are utilized, each of which consists of 10 irrelevant classes and each contains 100 images of the Corel stock photo database. The first subset is the Wang database that contains: Africa, Beach, Building, Bus, Dinosaur, Elephant, Flower, Horse, Mountain and Food classes. While the second one is Corel-1000 that comprises; Dinosaur, Cyber, Horse, Bonsai, Texture, Fitness, Dishes, Antiques, Elephant and Easter egg groups. Fig. 2, Fig. 3, Fig. 4 and Fig. 5 illustrate samples of the utilized datasets.

These databases are selected to represent two different types of CBIR chores: The ALOI and COIL-100 datasets characterize retrieval task that involves depicting more clearly defined objects with various viewing angles while Wang and Corel-1000 databases signify retrieval task with real world arbitrary photographs.

B. Implementation Details

All assessments were accomplished on a Lenovo laptop with Intel Core i7, 2.20 GHz processor, 8GB RAM, and Windows 10 Home Ultimate 64-bit as an operating system. The system was implemented in Matlab R2013b.

In this work, two validation techniques are utilized, viz. repeated holdout validation and k-fold cross-validation. The repeated holdout method splits randomly the data into two disjoint subsets; training set and test set and repeats this process with different subsamples. On the other hand, the k-fold validation generates a k subsets of equal size. The system is trained with $k - 1$ subsets and the remaining one forms the test set. This procedure is repeated k times. The holdout is simpler and needs less computation, however, there are overlapping test sets. On the contrary, k-fold cross-validation has the advantage that there are non-overlapping test sets. All samples in the dataset are ultimately used for both training and testing. However, it is computationally expensive.

Holdout validation is achieved by randomly choosing 90% of a dataset images for training and the remaining 10% of the images are reserved for testing. This validation procedure is repeated five times and the average performance is computed. Besides, ten-fold cross-validation is employed to evaluate the system performance.

To evaluate the proposed retrieval system, 4-level curvelet transform is applied. Besides, different quantization levels are used to compute the color histogram (CH) so as to obtain the optimum number of bins. The tests are carried out on 5 quantization levels; 9, 16, 25, 64 and 100 bins for each of the RGB channels.



Figure 2. Sample images from ALOI database.



Figure 3 . Sample images from COIL-100 database.

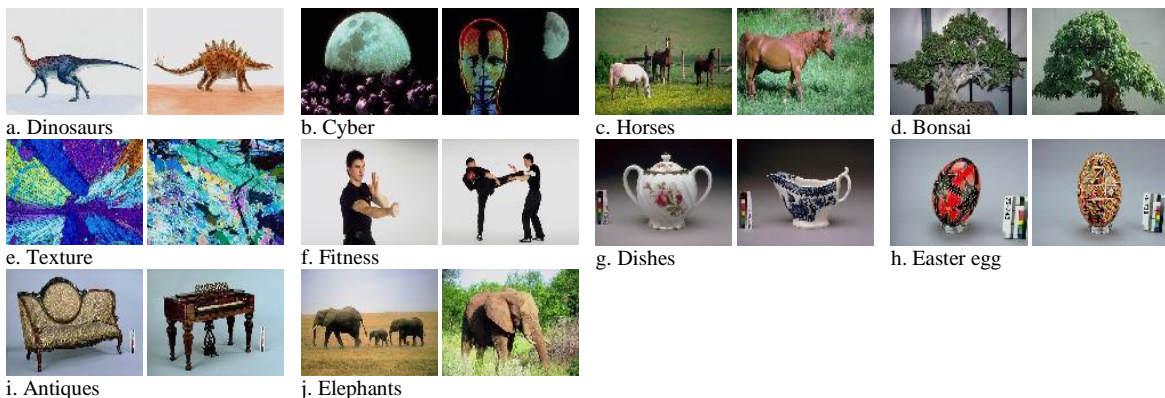


Figure 4. Sample images from Corel-1000 database.

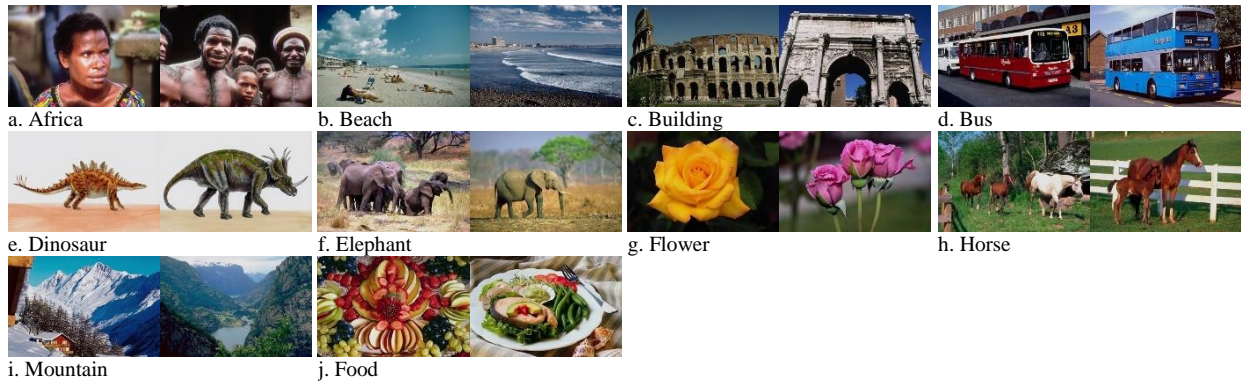


Figure 5. Sample images from Wang database.

Recall and precision metrics are utilized to measure the performance of the proposed CBIR system. The two metrics are defined as follows:

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \quad (4)$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieval}}{\text{Total number of relevant images in the database}} \quad (5)$$

Furthermore, Precision-Recall curve (PRC) is used to assess the effectiveness of the proposed image retrieval system.

C. Experimental results and discussion

In this section, the author presents and debates the results of the experimental evaluation of the proposed retrieval system. To estimate the efficiency of all implemented features (curvelet descriptors, quantized RGB CH (QCH) and CEDD) on the given databases, the author extracted these features from the images and executed experiments to test the effect of curvelet descriptors and CEDD features individually, the combination of both as well as the combination of them with the QCH. In addition, the author scrutinized the effect of using different quantization levels (five quantization levels are studied; 9, 16, 25, 64, 100). For all experiments, the author reports the average precision and recall ratios for the holdout and K-fold validation methods. Moreover, the results are benchmarked with previous works that use the same databases utilized in this work. Also, the author has compared the results of her previous work [30] with that of this work.

In this work, each image is represented by a feature vector of size

$$\begin{aligned} \text{feature vector size} &= N_{\text{curvelet}} + N_{\text{CEDD}} + N_{\text{CH}} \\ N_{\text{CH}} &= i \times 3 \end{aligned} \quad (6)$$

where N_{curvelet} is the dimension of curvelet descriptor (52 for 4-level decomposition), N_{CEDD} is the dimension of CEDD vector (144), N_{CH} is the total number of bins of the 3D-color histogram and i is the number of bins for each RGB channel (since RGB channels are equally quantized). Table 2 depicts the dimensions of the extracted feature vectors for the different realized methods.

Fig. 6 and Fig. 7 represent the experiments conducted on ALOI dataset using different retrieval methods. The results indicate that joining CEDD and curvelet improves the retrieval system significantly and combining them

with color histogram descriptors further enhances the retrieval performance. The optimal precision using K-fold cross-validation (0.995) and using holdout validation (0.996) is attained when the quantization level is 16 and 100 for each RGB channel, whereas it is near optimal at quantization levels 9, 25 and 64, with tiny variations. While the best recall (0.994) is obtained with 25 quantization levels using holdout validation as well as 100 quantization levels using K-fold validation. The proposed system achieved a good performance during both K-fold and holdout validations. This demonstrates the efficiency of the proposed retrieval system.

Fig. 8 and Fig. 9 summarize the experimental results of all implemented descriptors on the COIL-100 dataset. The results reported in Fig. 8 indicate that the best precision ratio (0.998) is reached by integrating CEDD, curvelet and RGB color histogram with 25 bins using holdout validation and with 9 and 25 bins using K-fold validation. As well, it is nearby to optimum (0.997), with an insignificant difference, when using 16 bins for both methods of validation. Additionally, Fig. 9 illustrate that the best recall ratio (0.998) is attained by joining CEDD, curvelet and RGB color histogram with 9 and 25 bins employing K-fold validation, and it is near to optimal (0.997) using 16, 25 and 64 bins using holdout validation as well as 64 bins using K-fold validation. It is also worth to be noted that recall close to optimum (0.997) when joining CEDD and curvelet descriptors for the holdout validation.

Fig. 10 and Fig. 11 report the results obtained by employing the proposed retrieval techniques on the Corel-1000 dataset. The author realized from these figures that integrating CEDD and curvelet descriptors enhances the precision ratio for both validation methods, the precision reaches its maximum value (0.954) using K-fold validation. Moreover, merging these descriptors with RGB color histogram with 25 bins further improves the precision value (0.964), and it is close to optimal (0.955) with 16 bins using holdout validation as well as (0.951) with 9 bins using K-fold validation. Also, utilizing 16 and 25 bins yield to almost the best recall ratio (0.948 and 0.943 using K-fold and holdout validation, respectively).

Fig. 12 and Fig. 13 present the comparison results of the Wang dataset using the proposed retrieval methods. From the results, the author recognized that the optimal precision ratio is reached utilizing CEDD, curvelet and RGB color histogram with 16 bins (0.898) using holdout validation and 9 bins (0.892) using K-fold validation with a tiny difference (0.006). Furthermore, recall ratio achieves the best results when employing CEDD, curvelet and RGB color histogram with 16 bins (0.874 using holdout validation) and 9 bins (0.877 using K-fold validation).

It can be noticed from the results that CBIR using the CEDD has a better performance than using curvelet this is because CEDD has color and texture features. Moreover, joining both descriptors enhances the performance significantly. Besides, integrating QCH descriptor improves the retrieval efficiency further and this improvement varies as the quantization levels vary. The author can also perceive that the use of greater quantization levels does not necessarily lead to a better precision. Contrarily, it leads to a much less efficient search. In CBIR, the retrieval effectiveness is essential than slight precision gain. Small retrieval performance enhancement at the cost of much higher dimension will reduce the entire system efficiency. Thus, for retrieval efficiency, the author recommends that CEDD, curvelet and RGB color histogram obtained from either 9, 16 or 25 quantization levels are appropriate to accomplish the retrieval tasks for all stated datasets.

TABLE 2
THE FEATURE VECTOR SIZE FOR THE IMPLEMENTED RETRIEVAL METHODS

Method	Feature vector size
CEDD	144
Curvelet	52
CEDD + Curvelet	196
CEDD + Curvelet + QCH with 9 bins	223
CEDD + Curvelet + QCH with 16 bins	244
CEDD + Curvelet + QCH with 25 bins	271
CEDD + Curvelet + QCH with 64 bins	388
CEDD + Curvelet + QCH with 100 bins	496

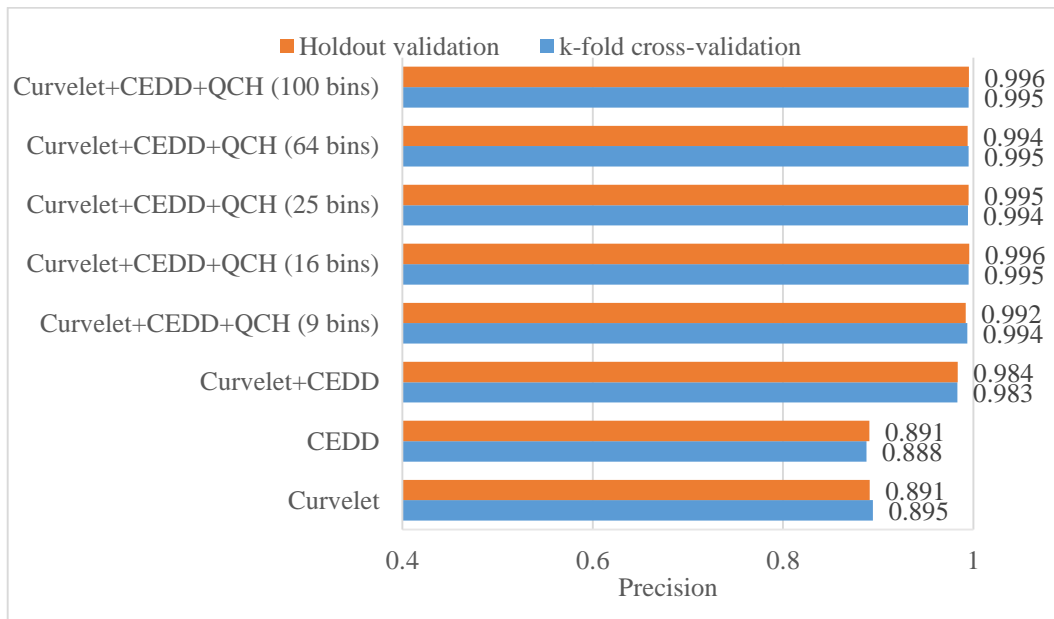


Figure 6. Average precision for different methods on ALOI dataset

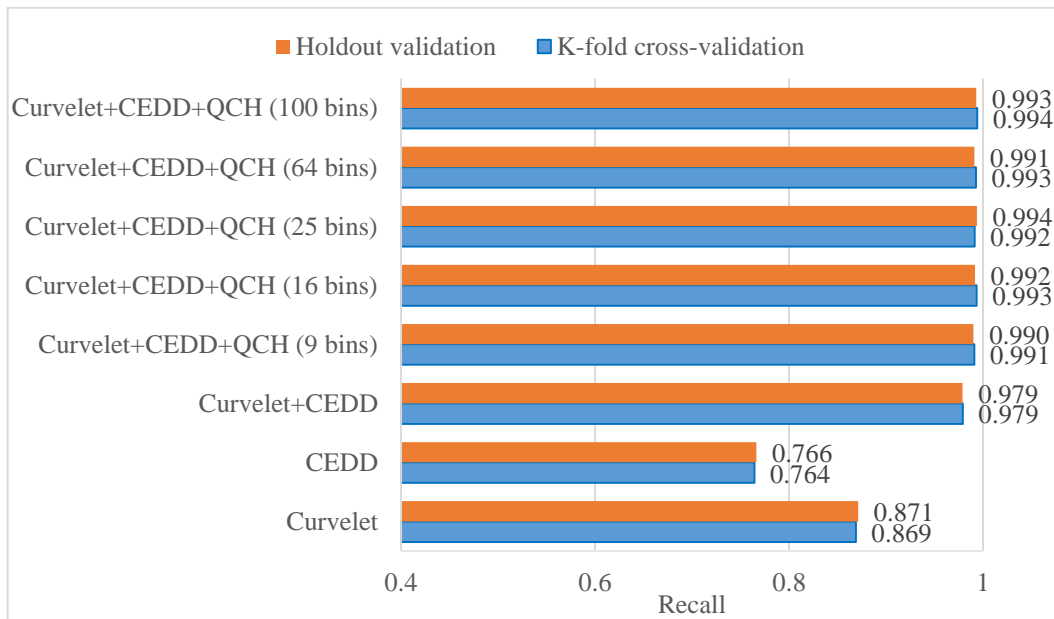


Figure 7. Average recall for different methods on ALOI dataset

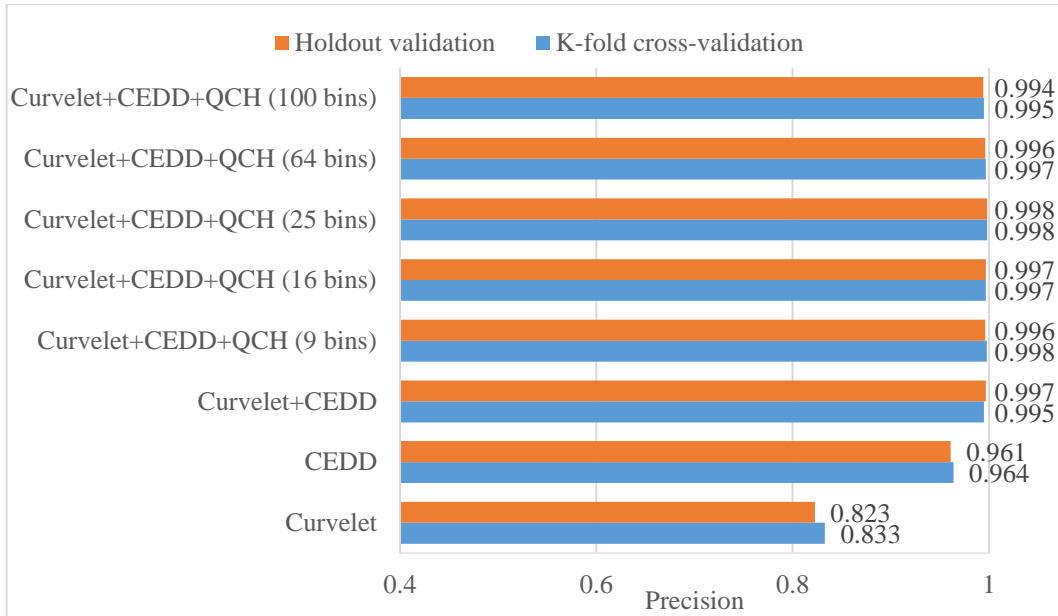


Figure 8. Average precision for different methods on COIL-100 dataset

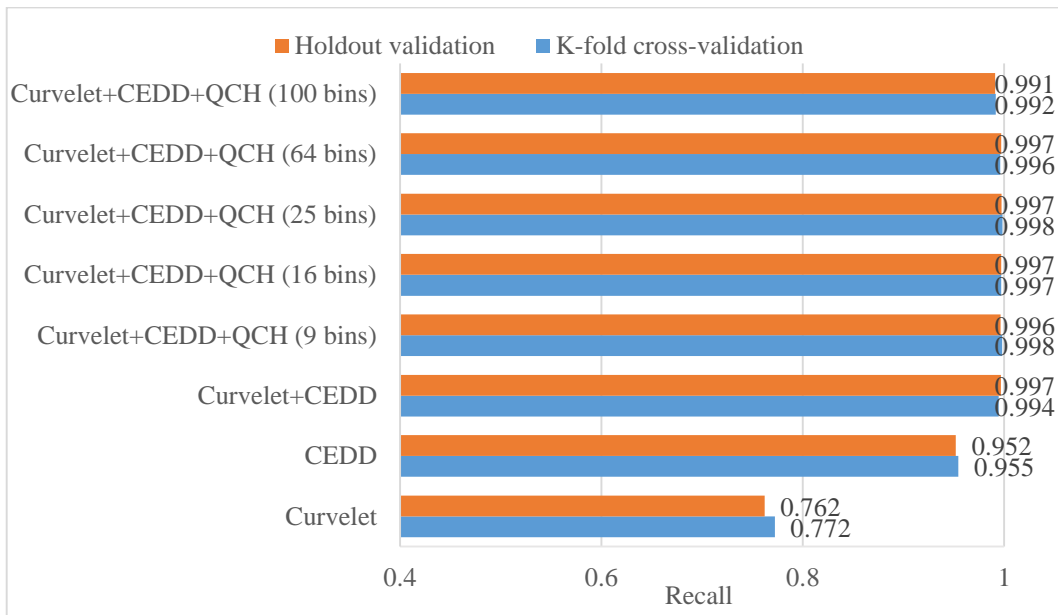


Figure 9. Average recall for different methods on COIL-100 dataset

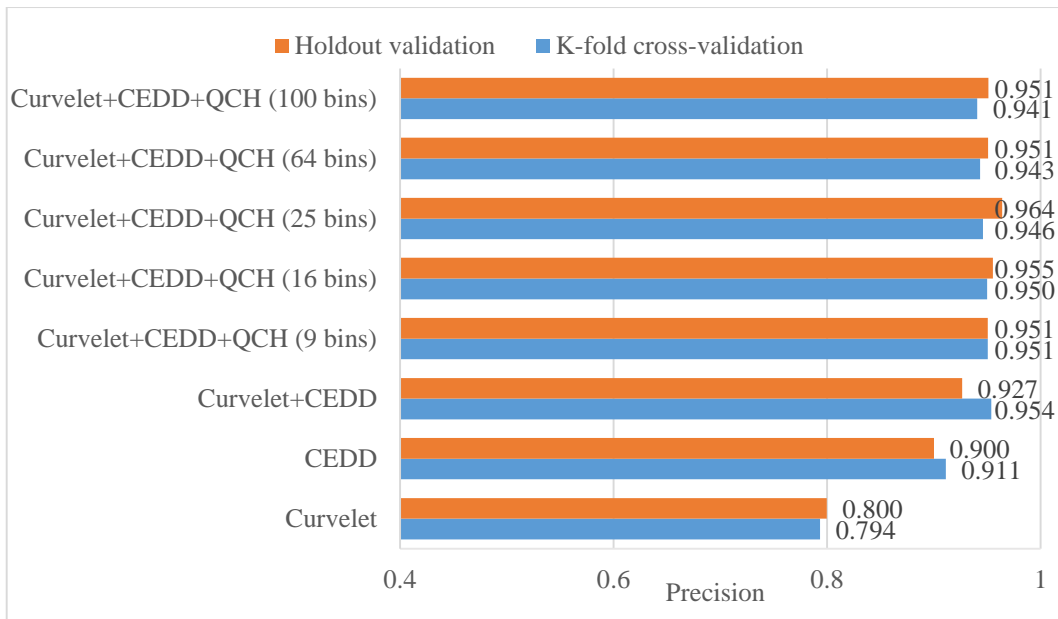


Figure 10. Average precision for different methods on Corel-1000 dataset

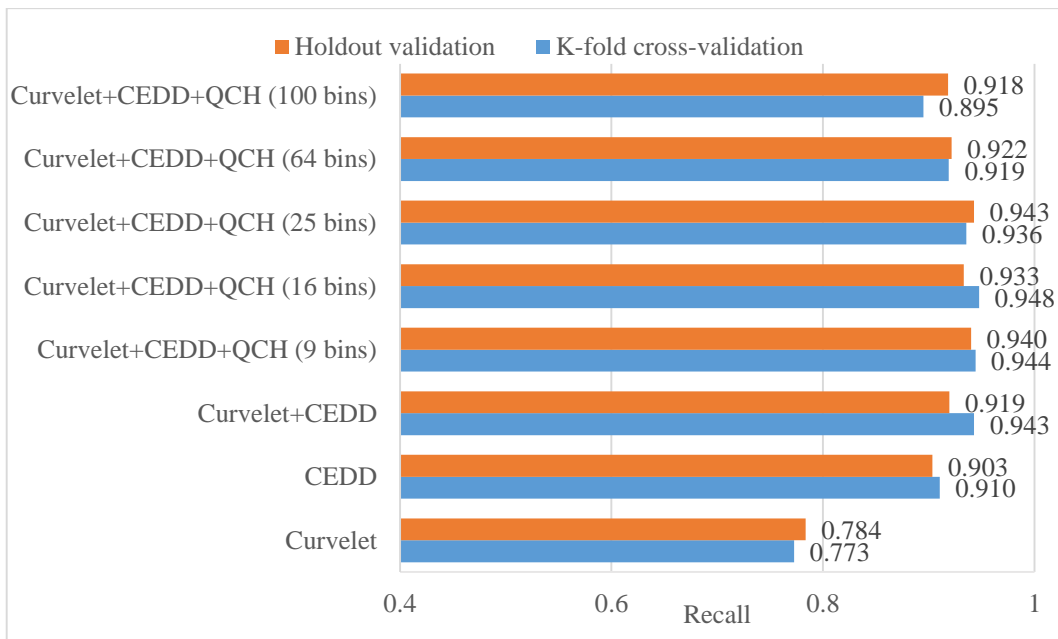


Figure 11. Average recall for different methods on Corel-1000 dataset

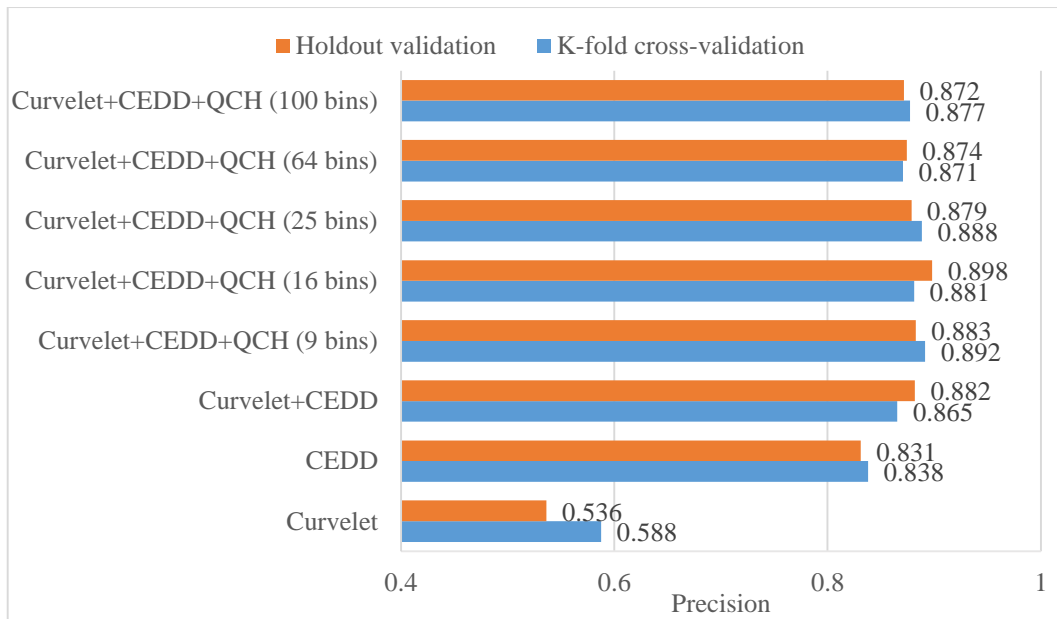


Figure 12. Average precision for different methods on Wang dataset

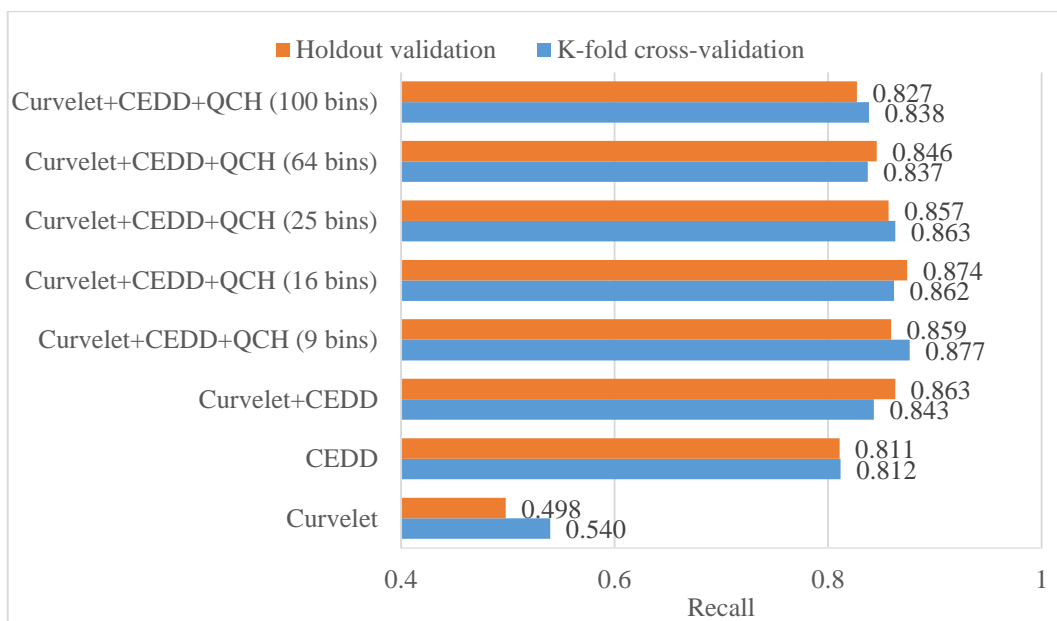


Figure 13. Average recall for different methods on Wang dataset

Moreover, the author has realized from the results that different features perform differently on the various datasets. Fig. 14 and Fig. 15 present the average precision of applying CEDD, curvelet and RGB color histogram with 9, 16 and 25 bins descriptors over all databases using K-fold and holdout validations, respectively. It is obvious that the system performs very well on ALOI and COIL-100 datasets compared to the Wang and Corel-1000 datasets, this is mainly because of the consistent background in all images in case of ALOI and COIL-100 datasets.

Furthermore, the author compared her retrieval method with other existing techniques. Table 3, table 4, table 5 and table 6 report comparisons between the proposed method and a group of other techniques on the explored

datasets. The performance of the proposed system, evaluated in terms of precision, achieved good results and the proposed model is competent with all the compared models.

The results obtained on ALOI and Corel-1000 datasets denote that the proposed system significantly outperforms previously addressed methods. However, for COIL-100 dataset, the method suggested in [35] shows comparable retrieval precision. Nevertheless, this method examines only 10 objects that were selected randomly from the whole dataset.

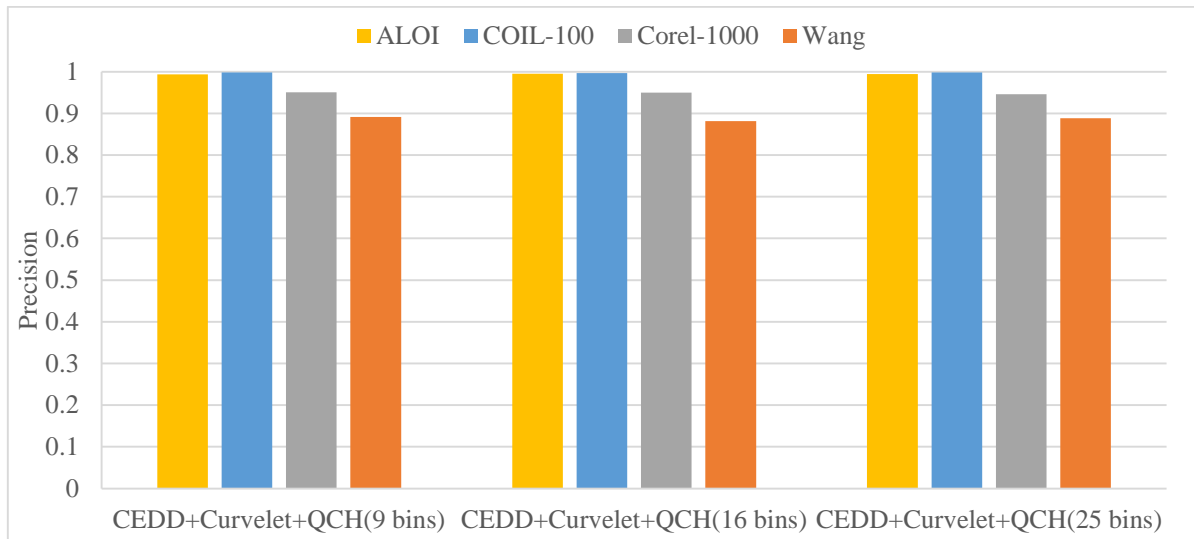


Figure 144. Average precision using K-fold validation different extracted features over each database

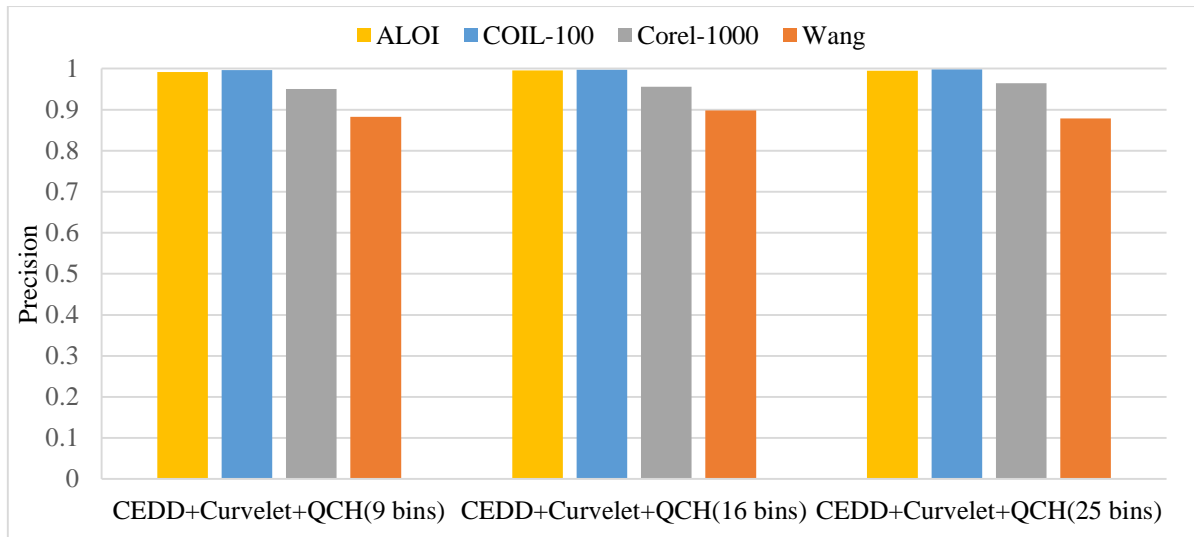


Figure 155. Average precision using holdout validation for different extracted features over each database

TABLE 3
COMPARISON WITH OTHER METHODS ON COIL-100 DATASET

Method	Precision
The proposed method (with 25 bins)	0.998
Kavya and Shashirekha [31] (10 random objects were only considered)	0.86
Velmurugan and Baboo [19] (15 random objects were only considered)	0.88
Mukherjee et al. [20] (10 random objects were only considered)	0.86
Bahri and Zouaki [32] (15 random objects were only considered)	0.78
Elnemr [30]	0.93
Khatabi et al. [35] (10 random objects were only considered)	0.9985

TABLE 4
COMPARISON WITH OTHER METHODS ON ALOI DATASET

Method	Precision
The proposed method (with 25 bins)	0.995
Alkhwilani, Elmogy and Elbakry [13] (10 random objects were only considered)	0.88

TABLE 5
COMPARISON WITH OTHER METHODS ON WANG DATASET

Method	Precision
The proposed method (with 16 bins)	0.898
Ali et al. [14]	0.877
Iakovidou et al. [15]	0.82
Rashno, Sadri and SadeghianNejad [17]	0.63
Mehmood et al. [33]	0.84
Vilvanathan and Rangaswamy [34]	0.75

TABLE 6
COMPARISON WITH OTHER METHODS ON COREL-1000 DATASET

Method	Precision
The proposed method (with 4 bins)	0.95
Elnemr [30]	0.88

V. CONCLUSION

This paper proposes a new CBIR technique that is based on integrating CEDD, curvelet and QCH descriptors. The classification stage is performed using a multiclass SVM. Generally, the precision of a CBIR system decreases as the number and variety of images increases in the dataset. Thus, the author assessed her proposed retrieval technique on benchmark databases from various domains such as to cover a wide range of different CBIR applications. The performance analysis is evaluated by computing the precision and recall as metrics. K-fold cross-validation and holdout validation are used to validate the results of implementing the various investigated descriptors as well as the different quantization levels. The experimental results are analyzed on the basis of comparing the retrieval performance of CEDD and curvelet individually and jointly as well as integrating CEDD, curvelet, and different QCH bins. The results indicate that merging CEDD and curvelet descriptors enhance the retrieval performance significantly. Furthermore, integrating them with QCH improves the performance more. The author also concludes that combining CEDD, curvelet and RGB color histogram with 9, 16 or 25 bins descriptors outperformed other examined descriptors regarding efficiency and scalability. Additionally, comparisons with existing CBIR techniques illustrate the effectiveness and efficiency of the proposed method.

Thus, the author hereby demonstrates the prospect of creating a better CBIR system with more significant feature sets.

In future, the employed datasets can be increased and new classes can be added to design a generalized and efficient retrieval system. Furthermore, advanced techniques like deep learning may be used to develop an efficient system for image retrieval and annotation.

REFERENCES

- [1] M. Yasmin, S. Mohsin, and M. Sharif, "Intelligent image retrieval techniques: A survey," *Journal of Applied Research and Technology*, vol. 12, no. 1, pp. 87-103, Feb. 2014.
- [2] K. Jenni, S. Mandala, and M. Sunar, "Content based image retrieval using colour strings comparison," *Procedia Computer Science*, vol. 50, pp. 374-379, 2015.
- [3] H. Elnemr, N. Zayed, and M. Fakhreldin, "Feature extraction techniques: fundamental concepts and survey," in *Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing*, N. Kamila, Ed. IGI Global, 2016, pp. 264-294.
- [4] S. Brandt, J. Laaksonen, and E. Oja, "Statistical shape features for content-based image retrieval," *Journal of Mathematical Imaging and Vision*, vol. 17, no. 2, pp. 187-198, Sep. 2002.
- [5] S. Gandhani, and N. Singhal, "Content based image retrieval: survey and comparison of CBIR system based on combined features," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 10, pp. 155-162, Oct. 2015.
- [6] K. Meskaldji, S. Boucherkha, and S. Chikhi, "Color quantization and its impact on the color histogram based image retrieval accuracy," in *NDT '09 First International Conf. on Networked Digital Technologies*, 2009, pp. 515-517.
- [7] S. Chatzichristofis, and Y. Boutalis, "CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval," in *Computer Vision Systems*, A. Gasteratos, M. Vincze and J. Tsotsos, Ed. Berlin Heidelberg: Springer-Verlag, 2008, pp. 312-322.
- [8] J. Starck, E. Candès and D. Donoho, "The Curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670-684, Jun 2002.
- [9] M. Angshul, "Bangla basic character recognition using digital Curvelet transform," *Journal of Pattern Recognition Research*, vol. 2, no. 1, pp. 17-26, 2007.
- [10] G. Joutel, V. Eglin, S. Bres, and H. Emptoz, "Curvelet based feature extraction of handwritten shapes for ancient manuscripts classification," in *Proc. of SPIE-IS&T Electronic Imaging*, Jan 2007, pp. 1-12.
- [11] Rajakumar, and Muttan, "A framework for MRI image retrieval using Curvelet transform and Euclidean distance," *Journal of Computer Science*, vol. 9, no. 3, pp. 285-290, 2013.
- [12] I. J. Sumana, M. Islam, D. Zhang, and G. Lu, "Content based image retrieval using curvelet transform," in *IEEE 10th workshop on Multimedia signal processing*, Australia, Oct. 2008, pp. 11-16.
- [13] M. Alkhwilani, M. Elmogy and H. Elbakry, "Content-based image retrieval using local features descriptors and Bag-of-Visual Words," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 6, no. 9, pp. 212-219, 2015.
- [14] N. Ali, K. B. Bajwa, R. Sablatnig, and Z. Mehmood, "Image retrieval by addition of spatial information based on histograms of triangular regions," *Computers and Electrical Engineering*, vol. 54, pp. 539-550, Aug. 2016.
- [15] C. Iakovidou, N. Anagnostopoulos, A. Kapoutsis, Y. Boutalis, M. Lux, and S. Chatzichristofis, "Localizing global descriptors for content-based image retrieval," *EURASIP Journal on Advances in Signal Processing*, vol. 80, pp. 1-20, Dec. 2015.
- [16] F. Malik, and B. Baharudin, "Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain," *Journal of King Saud University – Computer and Information Sciences*, vol. 25, no. 2, pp. 207-218, July 2013.
- [17] A. Rashno, S. Sadri, and H. SadeghianNejad, "An efficient content-based image retrieval with ant colony optimization feature selection schema based on wavelet and color features," in *International Symposium on Artificial Intelligence and Signal Processing (AISP)*, Iran, March 2015, pp. 59-64.
- [18] E. Walia, and A. Pal, "Fusion framework for effective color image retrieval," *Journal of Visual Communication and Image Representation*, vol. 25, no. 6, pp. 1335-1348, Aug. 2014.
- [19] K. Velmurugan, and S. Baboo, "Content-based image retrieval using SURF and color moments," *Global Journal of Computer Science and Technology*, vol. 11, no. 10, pp. 1-5, May 2011.
- [20] A. Mukherjee, S. Chakraborty, J. Sil, and A. Chowdhury, "A novel visual word assignment model for Content-Based Image Retrieval," in *Proceedings of International Conference on Computer Vision and Image Processing, Advances in Intelligent Systems and Computing*, Springer, Singapore, 2017, pp. 79-87.
- [21] K. Jayanthi and M. Karthikeyan, "Color and edge directive descriptor feature extraction technique for content based image retrieval system," *Middle-East Journal of Scientific Research*, vol. 23, no. 8, pp. 1590-1597, 2015.
- [22] S. Chatzichristofis, and Y. Boutalis, "A hybrid scheme for fast and accurate image retrieval based on color descriptors," in *IASTED International Conference on Artificial Intelligence and Soft Computing (ASC 2007)*, Spain, Aug. 2007, pp. 280-285.
- [23] E. Candès, and D. Donoho, "Curvelets - a surprisingly effective nonadaptive representation for objects with edges," Department of Statistics, Stanford University, Report Number 1999-28, Nov. 1999.
- [24] J. Starck, E. Candès and D. Donoho, "The Curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670-684, Aug. 2002.
- [25] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete Curvelet transforms," *Multiscale Modeling and Simulation*, vol. 5, no. 3, pp. 861-899, March 2006.
- [26] S. Arivazhagan, L. Ganesan, and T. Kumar, "Texture classification using Curvelet statistical and co-occurrence features," presented at the 18th International Conference on Pattern Recognition (ICPR'06), Aug. 20-24, 2006, China.
- [27] J. Geusebroek, G. Burghouts, and A. Smeulders, "The Amsterdam library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103-112, 2005.
- [28] S. Nene, S. Nayar, and H. Murase, "Columbia object image library (COIL-100)," Department of Computer Science, Columbia University, Technical Report CUCS-006-96, March 1996.
- [29] J. Wang, J. Li, and G. Wiederhold, "SIMPLcity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 23, no. 9, pp. 947-963, Sep. 2001.
- [30] H. Elnemr, "Combining SURF and MSER along with color features for image retrieval system based on bag of visual words," *Journal of Computer Science*, vol. 12, no. 4, pp. 213-222, 2016.

- [31] J. Kavya, and H. Shashirekha, "A Novel approach for image retrieval using combination of features," *International Journal of Computer Technology & Applications*, vol. 6, no. 2, pp. 323-327, Jan. 2015.
- [32] A. Bahri, and H. Zouaki, "A SURF-Color moments for images retrieval based on bag-of features," *European Journal of Computer Science and Information Technology*, vol. 1, no. 1, pp. 11-22, Jun 2013.
- [33] Z. Mehmood, S. Anwar, N. Ali, H. Habib, and M. Rashid, "A novel image retrieval based on a combination of local and global histograms of visual words," *Mathematical Problems in Engineering*, vol. 2016, pp. 1-12, Jun 2016.
- [34] K. Vilvanathan, and R. Rangaswamy, "Bi-Level classification of color indexed image histograms for content based image retrieval," *Journal of Computer Science*, vol. 9, no. 3, pp. 343-349, 2013.
- [35] Khatabi, A. Tmiri and A. Serhir, "An efficient method of improving image retrieval using combined global and local features," *Advances in Ubiquitous Networking 2. Lecture Notes in Electrical Engineering*, Springer, Singapore, 2017, pp. 431-443.

A survey on multi-objective task scheduling algorithm in cloud environment

Tanvi Gupta

Dr.SS.Handa

Dr. Supriya Panda

Research Scholar

Professor

Professor

Manav Rachna International Institute of Research and Studies SRM University Manav Rachna International Institute of Research and Studies

Abstract: Cloud computing is one of the important subject now- a- days in which services are given to the users by cloud provider. So, according to the use of the services users will pay to the providers. Resource allocation and task scheduling are important to manage the task in cloud environment for load balancing. Task scheduling is an important step to improve the overall performance of the cloud computing. Task scheduling is also essential to reduce power consumption and improve the profit of service providers by reducing processing time. So, for task scheduling, various “quality of service” (QoS) parameters are considered for reducing execution time and maximize throughput. For this purpose, multi-objective optimization of task scheduling is used, which is a sub domain of “multi-criteria decision making” problem. This involves more than one objective function that can be optimized at the same time.

Keyword: Cloud computing, multi-objective, QoS parameter

Introduction

In the IT industry, Cloud computing [5] is the latest buzzword. With the foundations of grid computing, utility computing, service oriented architecture, virtualization and web 2.0; it is an emerging computing paradigm. With the ownership of just an Internet connection, the user can access all the required software, hardware, platform, applications, infrastructure and storage. A cloud is a type of parallel and distributed system a collection of interconnected and virtualized computer that are dynamically provisioned and presented as one or more unified computing resources based on SLAs established through negotiation between the service providers and consumers. In this information technology oriented growing market of businesses and organizations, cloud computing provides virtual resources that are dynamically scalable. It describes virtualized resources, software, platforms, applications, computations and storage to be scalable and provided to users instantly on payment for only what they use [5].

Multi-objective Optimization

Optimization [5] deals with the problems of seeking solutions over a set of possible choices to optimize certain criteria. They become single objective optimization problems, if there is only one criterion to be taken into consideration, which

have been extensively studied for the past 50 years. So, we have multi-objective optimization problems, if there is more than one criterion which must be treated simultaneously. Multiple objective problems arise in the design, modeling and planning of complex real systems in area of industrial production, urban transportation, capital budgeting, forest management, reservoir management, layout and landscaping of new cities, energy distribution, etc. It is easy to see that almost every important real-world decision problems involves multiple and conflicting objectives, which need to be tackled while respecting various constraints, leading to overwhelming problem complexity. The multiple objective problems have been receiving growing interest from researchers with various backgrounds since early 1960. There are a number of scholars who have made significant contributions to the problem. Among them Pareto is perhaps one of the most recognized pioneers in the field.

Scheduling and scheduling criteria

Scheduling is to allocate task to appropriate machine to achieve some objectives or we can say it determines on which machine which task should be executed. In traditional scheduling tasks are directly mapped to resources at one level, whereas, now a days , resources in cloud are scheduled at two level i.e. physical level and VM level which is depicted in Figure1. There are mainly two types of task scheduling in cloud computing: static scheduling and dynamic scheduling. In static task scheduling, information of task is known before execution like execution time whereas in dynamic task scheduling, information of task is not known before execution [11]. In cloud environment to execute a task a user request for a computing resource which is allocate by cloud provider after finding the appropriate resource among existing as shown in Figure 1. Tasks which are submitted for execution by users may have different requirements like execution time, memory space, cost, data traffic, response time, etc. Also, the resources which are involved in cloud computing may be diverse and geographically dispersed. There are different environments in cloud: single cloud environment and multi-cloud environment.

Scheduling process [6] in cloud can be categorized into three stages namely–

a. Resource discovering and filtering – Resource request is made by cloud user and submitted to service provider; service provider searches the suitable resources to locate them.

b. Resource selection –Resource is selected on the basis of task and resource selection parameters.

c. Task submission -Task is submitted to the selected resource.

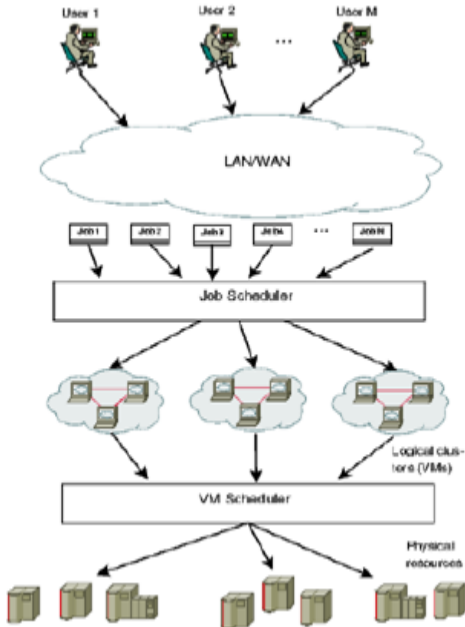


Figure1 :View of cloud

Scheduling criteria

The criteria differ with respect to service provider and user. Service provider wants to gain revenue, maximize resource utilization with minimal efforts, whereas, user wants his job to be executed with minimal cost in minimum time [10]. Figure 2 below shows the Conceptual modeling of the Cloud Computing environment for Task Scheduling.

A. Cloud User Preferred

1) **Makespan**: it tells about the finishing time of last task. The makespan should be minimum, which shows the fastest execution time of a task.

2) **Cost**: it is the sum of the amount paid by the user to provider for using individual resource.

3) **Waiting time**: the time spent by a task in ready queue to get a chance for execution.

4) **Turnaround time (TAT)**: Time taken by a task to complete its execution after its submission i.e., the sum of waiting time and execution time of a task.

5) **Tardiness**: the delay in execution of a task i.e. difference between finishing time and deadline of the task. For an optimal scheduling the tardiness should be zero which shows no delay in execution.

6) **Fairness**: this shows that all tasks are getting equal opportunity of execution.

7) **Response time**: time taken by a system to start responding (first response) after submission of a task.

B. Cloud provider preferred

1) **Resource utilization**: the resources should be fully utilized by keeping them as busy as possible to gain the maximum profit.

2) **Throughput**: his represents the number of task completed in a per time unit.

3) **Predictability**: this represents the consistency in the response times of task .Unpredictable response time may degrade the performance of system.

4) **Priority**: To give preference to a task to finish it as earliest. Priority can be given on the basis of arrival time, execution time, deadline etc. Resources are provided to higher priority task to complete the execution.

5) **Load balancing**: distribution of load among all the computing resources.

6) **Deadline**: the time till which a task should be completed.

7) **Energy efficiency**: Reducing the amount of energy used to provide any solution or service.

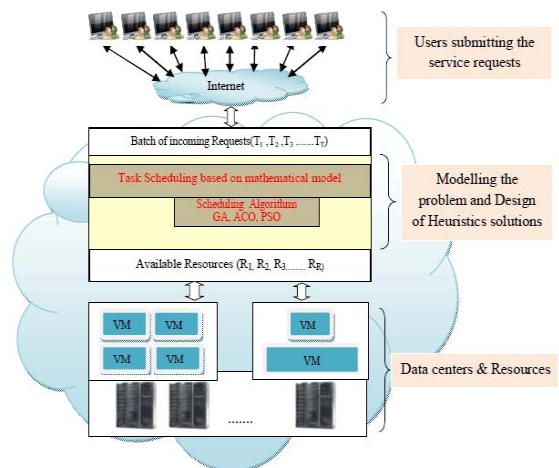


Figure 2 : Conceptual modeling of the Cloud Computing environment for Task Scheduling

Literature Survey

Dynamic Multi-objective task scheduling in Cloud Computing based on Modified particle swarm optimization(2015): A.I.Awad et.al[1] state the efficient allocation of tasks to available virtual machine in user level base on different parameters such as reliability, time, cost and load balancing of virtual machine. Agent used to create dynamic system. The proposed mathematical model multi-objective Load Balancing Mutation particle swarm optimization (MLBMPSO) is used to schedule and allocate tasks to resource which is

shown below in Figure 3. MLBMP SO considers two objective functions to minimize round trip time and total cost. Reliability can be achieved in system by getting task failure to allocate and reschedule with available resource based on load of virtual machine.

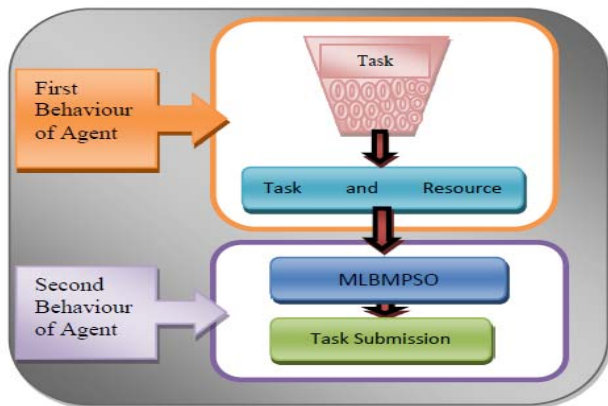


Figure 3: Proposed model by author

Figure 3[1] depicts two behaviors:
 First behaviour is responsible for
 1. Task Buffer
 2. Task and Resource Information
 Second behaviour is responsible for
 1. MLBMP SO
 2. Task Submission

Responsive Multi-objective Load Balancing Transformation Using Particle Swarm Optimization in Cloud Environment (2016): VG.Ravindhren.et.al [9] states that Resource allocation among multiple clients has to be ensured as per SLAs. So, to accomplish the goals and achieve high performance, it is important to design and develop a Responsive multi-objective load balancing Transformation algorithm (RMOLBT) based on abstraction in multi cloud environment. The model is represented below in Figure 4. It is the most challenging to schedule the tasks along with satisfying the user's Quality of Service (QoS) requirements. This paper proposes a wide variety of task scheduling and resource utilization using Particle swarm Optimization (PSO) in cloud environment. The results in this paper demonstrate the suitability of the proposed scheme that will increase throughput, reduce waiting time, reduction in missed process considerably and balances load among the physical machines in a Data centre in multi cloud environment.

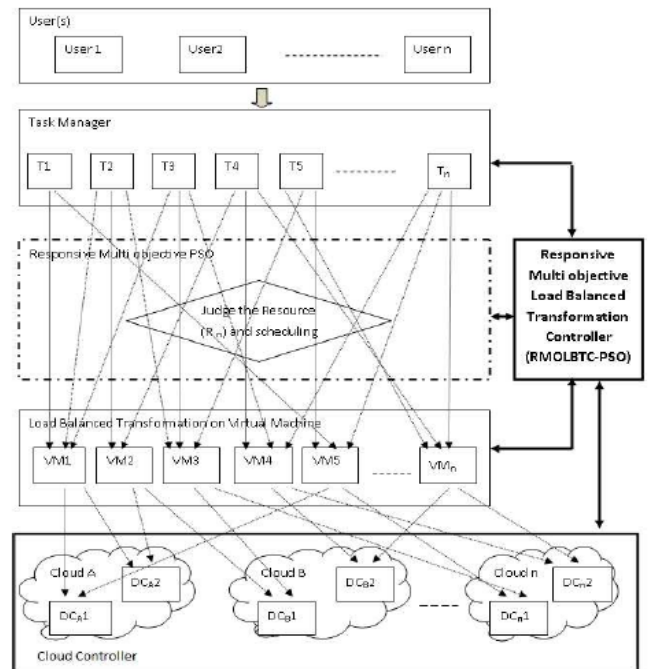


Figure 4: System Model for Responsive Multi objective Load Balance Transformation using PSO

An Efficient Approach for Task Scheduling Based on Multi-Objective Genetic Algorithm in Cloud Computing Environment (2014): Sourabh Budhiraja et.al[5] state that the scheduling of the cloud services to the consumers by service providers influences the cost benefit of this computing paradigm. In such a scenario, tasks should be scheduled efficiently such that the execution cost and time can be reduced. In this paper, the author proposed an efficient approach for task scheduling based on Multi- Objective Genetic Algorithm (MOGA) shown below in Figure 5, which minimizes execution time and execution cost as well. For task scheduling, a Multi-Objective genetic algorithm [5] is implemented and the research is focused on crossover operators, mutation operators, selection operators and the Pareto solutions method.

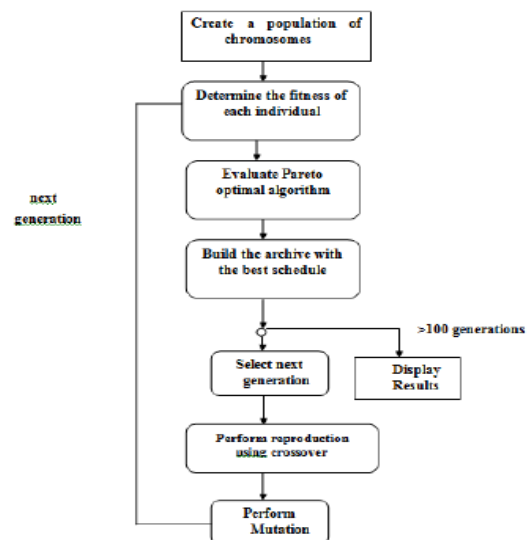


Figure5: Flow Chart-Modified Genetic Algorithm(MGA)[1]

Multi-Objective Tasks Scheduling Algorithm for Cloud Computing Throughput Optimization(2015): Atul Vikas Lakra et.al[2] state that the problem is to bind set of tasks received by the broker to the received list of VMs, so that execution time of workload is reduced to minimal optimized time. Single objective scheduling algorithms have some problem. For example, in priority task scheduling, high priority tasks always get chance to execute, due to this low priority task have to wait for a long time. Sometimes low priority task gets a chance to execute but, if high priority tasks keep coming then low priority task is pre-empted and CPU is allocated to high priority task and this leads to increase in execution time of a task as well as it reduces the throughput of the system. Similarly, in First Come First Serve (FCFS) and Shortest Job First(SJF) ,task scheduling algorithms face problem in worst case scenario. These algorithms perform very well in the best case but in worst case they degrade the performance to very low level. So an efficient scheduling algorithm is required which can provide optimized performance in both cases. Using a proper scheduling algorithm implementation in broker improves the datacenter’s performance without violating service level agreements. The order of task submission and the VMs also influence the execution time of the entire workload.

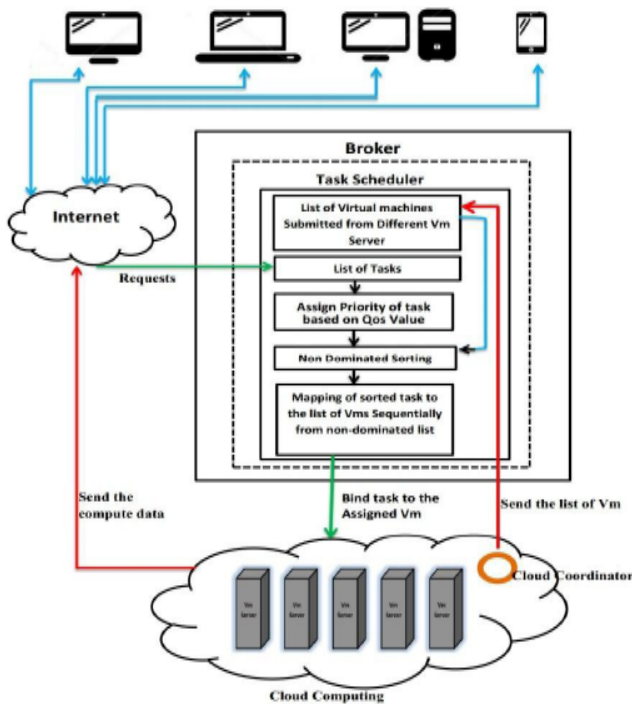


Figure6: Cloud computing architecture

The cloud computing architecture and the proposed work is shown above in Figure 6, which depicts that the Cloud broker is responsible for mediating negotiations between SaaS and cloud provider and such negotiation are driven by QoS requirements. Broker acts on behalf of SaaS for

allocation of resources that can meet application’s QoS requirements. Also, Figure 7 below represents the algorithm of multi –objective task scheduling [5] proposed by the author.

Algorithm 1 : Multi-objective task scheduling algorithm

```

1. Submit both VMs list of successfully created VMs in datacenter and task list to Broker.
2. Create a received list of tasks.
3. Create a received list of VMs.
4. Non-dominated sorting (list of task)
   i ← 0
   Create empty non-dominated list
   dominated list ← list of task
   Initially put taski in the non-dominated list
   for all i ← 1 to size of task's list do
     for all j ← 0 to size of non-dominated list do
       if taskj dominates taski then
         put taskj into non dominated set
       else
         if taski dominates taskj then
           put taski into non dominated set
         end if
       else
         put taski and taskj into non dominated set
       end if
     end for
   end for
5. Sort the list of task according to the non-dominated task set.
6. Sort the VM received list in descending order .
7. j ← 0.
   for all i ← 0 to the size of task's list do
     if j ≥ 0 then
       Bind taski to the VMj j++
       if j == number of VMs then
         j=0
       end if
     end if
   end for

```

Figure 7: Algorithm for multi-objective task scheduling[2]

Multi-Target Tasks Scheduling Algorithm for Cloud-environment Throughput Optimization(2016): Shubhashree S. R et.al[7]

proposed that proposed multi-task scheduling algorithm that enhances the data center execution without damaging SLA. The proposed algorithm is as appeared below in Figure 8 and Figure 9, that utilize non-dominating sorting algorithm for comprehending the multi-objective (task size, QOS value). After a fixed time, interval, the list will be updated dynamically. This algorithm will give the optimized throughput when compared with the existing algorithm.

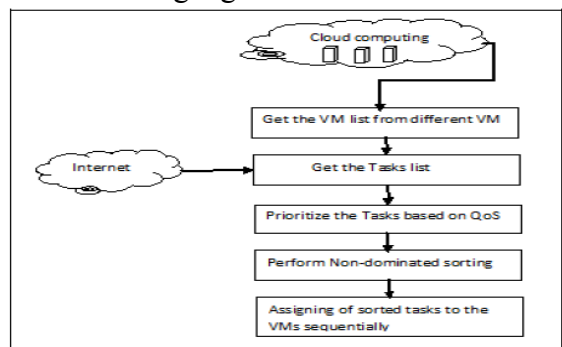


Figure8 :Multiobjective task scheduling and dominance relation

```

1. Submit both VMs list of successfully created VMs in datacenter and t
Broker.
2. Create a received list of tasks.
3. Create a received list of VMs.
4. Non-dominated sorting (list of task)
   i ← 0
   Create empty non-dominated list
   dominated list ← list of task
   Initially put taski in the non-dominated list
   for all i ← 1 to size of task's list do
     for all j ← 0 to size of non-dominated list do
       if taskj dominates taski then
         put taskj into non dominated set
       else
         if taski dominates taskj then
           put taski into non dominated set
         end if
       else
         put taski and taskj into non dominated set
       end if
     end for
   end for
5. Sort the list of task according to the non-dominated task set.
6. Sort the VM received list in descending order .
7. j ← 0
   for all i ← 0 to the size of task's list do
     if j ≥ 0 then
       Bind taski to the VMj; j++
       if j== number of VMs then
         j=0
       end if
     end if
   end for

```

Figure 9: Algorithm for multi-target task scheduling algorithm[7]

Multi-Objective Task Scheduling in Cloud Computing Using an Imperialist Competitive Algorithm(2016): Majid Habibi.et.al[3] states that Since the tasks scheduling in the cloud computing environment and distributed systems is an NP-hard problem, in most cases to optimize the scheduling issues, the meta-heuristic methods inspired by nature are used rather than traditional or greedy methods. One of the most powerful meta-heuristic methods of optimization in the complex problems is an Imperialist Competitive Algorithm (ICA). Thus, in this paper, a meta-heuristic method based on ICA is provided to optimize the scheduling issue in the cloud environment. Figure 10 below depicts the pseudo code proposed by the author in this paper.

```

Input: npop(Population-size),problem-size,ep,α,β,pr
For i=1 to npop do
  Ciposition ← RandomPosition(problem-size)
  If i<=ep then
    EmpiresPopulation ← Ciposition
  Else
    Cw ← GetWorstSolution(EmpiresPopulation)
    If Cost(Ciposition) < Cost(Cwposition) then
      Replace(EmpiresPopulation,Ci,Cw)
    Else
      Ciempire ← assignAnEmpire(EmpiresPopulation)
    End
  End
  Populaton ← Ci
End
EvaluatePopulaton(Population)
EvaluateEmpiresPopulation(EmpiresPopulation,Population)
ImperialisticCompetition(EmpiresPoplution,Population)
EliminiatWeakestEmpire(EmpiresPoplution,Population)

```

```

End
EvaluatePopulation(Population)
BestSol ← GetBestSolution(Population)
Return BestSol

```

Figure 10: Pseudo code for the algorithm proposed by the author [3]

Multi-Objective Task Scheduling using K-mean Algorithm in Cloud Computing(2016):Vanita Dandhwani.et.al[8]

states that K-mean clustering algorithm is used to create the clusters for tasks. In which for k clusters centroids are calculated based on multi-objectives Task length and Deadline using equation (1) and (2) and Centroid is calculated using equation (3) where minimum distance value is selected as centroid.

$$Tl = \text{Number of Instructions (MI)} \quad (1)$$

$$Dl = \text{VMmips} / Tl \quad (2)$$

Where Tl=Tasksize

Dl=Deadline

VMmips= MIPS of Average VM

$$\text{dist}((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2} \quad (3)$$

Where x= tasksize

y=deadline

Figure 11 below shows the k-means multi objective task scheduling algorithm and resource selection algorithm

A. K-MEAN ALGORITHM

Step1: select k points as initial centroid.

Step2: Repeat

Step3: Form k cluster by assigning each point to its closest centroid.

Step4: Recompute the centroid for each cluster.

Step5: Untill centroid do not change.

B: MULTI-OBJECTIVE TASK SCHEDULING ALGORITHM

Step1: Get a list of unscheduled task.

Step2: Create a cluster using K-mean algorithm.

Step3: Arrange clusters in descending order (higher the centroid higher the cluster).

Step4: Map clusters to the Best VM using Resource selection algorithm.

C: RESOURCE SELECTION ALGORITHM

- Step1 : Input: Get Resource list.
- Step2: Begin i=0
- Step3: While cluster[i] is not empty do
- Step4: Select the VM which has maximum capacity using equation (4) [12].

$$C_i = \text{Pro}_{mi} * P_{mpi} + \text{VM}_{bwi} \quad (4)$$

Where Pro_{mi} is the number of processors in VM_i ,

P_{mpi} is millions of instructions per second of all processors in VM_i

VM_{bwi} is the communication bandwidth ability of VM_i

Step5: Schedule the cluster and execute it.

Step6: Update status of resources.

Step7: $i=i+1$

Step8: End while

Step9: End.

Figure 11: Algorithm proposed by the author [11]

Multi objective Task Scheduling in Cloud Environment Using Nested PSO Framework (2015): R K Jena.et.al[4] focuses on task scheduling using a multi-objective nested Particle Swarm Optimization(TSPSO) to optimize energy and processing time.

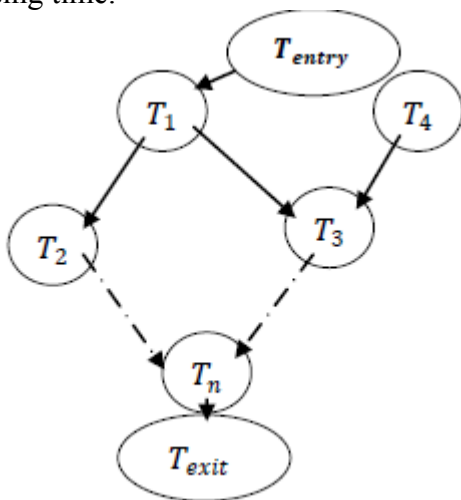


Figure12: Task graph

Figure12 above represents Each vertex V in the DAG is associated with a value $\langle l \rangle$, 'l' represents the length of the task in Million Instruction (MI). The problem of this model is how to optimally schedule user jobs to the Processing Elements available in the cloud under different data center. All the PEs is considered homogeneous, unrelated and parallel. Scheduling is considered as non preemptive, which means that the processing of any task can't be interrupted. Figure 13 represents the algorithm proposed by the author of this paper.

Algorithm MOPSO()

```

{
Initialize External Archive (AE) // AE = U E
For j = 1 to M (M is the size of particle swarm)
Initialize Sj & Vj // Initialization of each particle swarm and its velocity
For k=1 to L // L is the number of iteration
{
For j = 1 to M
E[j] = PSO (Sj) // E [j] is the archive for particle swarm Sj
Update the archive (AE) of non-dominated solutions

Select leader particle from the archive (E)//
Update velocity
Update position
}
Return ( Non-dominated solution)
}
  
```

Algorithm PSO(S_j)

```

{ // Sj . represents the set of user tasks of jth particle allocated to different data center (Dt),
t = 1...P
For t = 1 to P //P is the number of available data center in Cloud environment
{
For i = 1 to N (N is the size of particle swarm)
{
Initialize S[i]
Initialize the velocity V of each particle V[i]
Initialize the Personal Best pBest of each particle. pBest[i] = S[i]
Evaluate objectives of each particle: Evaluate S[i]
Initialize the Global Best particle (gBest) with the best one among the 'N' particles:
gBest = Best particle found in S
} // end of loop 'i'
Add the non-dominated solutions found in S into EA[t] // EA[t] is the External Archive storing the pareto front for
the task assign to data center Dt,
Initialize the iteration number (k) = 0
Repeat until k > G // (G is the maximum number of iterations)
{
For i = 1 to N (swarm size)
{
Randomly select the global best particle for S from the External Archive EA[t] and store its
position in gBest.
Calculate the new velocity V[i] according to (7)
Compute the new position of S[i] according to (8)
If (r < G * PMUT) then // (PMUT is the probability of mutation)
Perform mutation on S[i]
Evaluate S[i] using (2) and (3)
Update the personal best solution of each particle S[i]
Update the External Archive EA[t]
} // end of loop 'i'
}
Retain the best pareto solution in EA[t]
} // end of loop 't'
Return (Min { EA[1], Makespant=1..P }, Sum{ EA[1], Energyt=1..P })
}
  
```

Figure 13: Proposed algorithm by author [4]

Conclusion:

Above Literature summarizes the multi-objective task scheduling algorithm in one form or the other. As we know, single objective functions cannot fulfill all the criteria, e.g., if we consider only priority of the task, rest of the QoS factors, which are very important in scheduling are left like task length, execution time, deadline, cost etc. So, multi-objective task scheduling algorithm is important for enhancing the cloud environment performance. Below Table1 summarized the algorithms of multi-objective task scheduling proposed by the various authors.

Paper name	Authors' name	QoS parameters	Algorithm used	Which parameter is improved
Dynamic Multi-objective task scheduling in Cloud Computing based on Modified particle swarm optimization(2015)	A.I.Awad, N.A.El-Hefnawy and H.M.Abdel_kader	reliability, time, cost and load balancing of virtual machine(VM).	multi-objective Load Balancing Mutation particle swarm optimization (MLBMPSO) algorithm	Execution time and makespan is minimized
Responsive Multi-objective Load Balancing Transformation Using Particle Swarm Optimization in Cloud Environment(2016):	VG.Ravindhren and Dr. S. Ravimaran	Job size	PSO algorithm	increase throughput, reduce waiting time, reduction in missed process considerably and balances load among the physical machines in a Data centre
An Efficient Approach for Task Scheduling Based on Multi-Objective Genetic Algorithm in Cloud Computing Environment(2014):	Sourabh Budhiraja, Dr. Dheerendra Singh	Cost and size	Multi-Objective Genetic Algorithm (MOGA)	minimizes execution time and execution cost
Multi-Objective Tasks Scheduling Algorithm for Cloud Computing Throughput Optimization(2015)	Atul Vikas Lakraa, Dharmendra Kumar Yadav	Size, cost	Multi-objective task scheduling algorithm	better performance and improved throughput, reduced cost
Multi-Target Tasks Scheduling Algorithm for Cloud-environment Throughput Optimization(2016):	Shubhashree S. R	Size of task	<i>Non-Dominated Sorting algorithm</i>	<i>to enhance the data-center throughput, diminishes the execution time.</i>
Multi-Objective Task Scheduling in Cloud Computing Using an Imperialist Competitive Algorithm(2016):	Majid Habibi, Nima Jafari Navimipour	Task size	Imperialist Competitive Algorithm	Execution time
Multi objective Task Scheduling in Cloud Environment Using Nested PSO Framework	R K Jena	Size of task	multi-objective nested Particle Swarm Optimization(TSPSO)	optimize energy and processing time.
Multi-Objective Task Scheduling	Vanita Dandhwani, Dr.Vipul Vekariya	Task length and deadline	multi-objective task scheduling	minimize the execution time

using K-mean Algorithm in Cloud Computing(2016):			algorithm using k-mean clustering	and makespan
--	--	--	-----------------------------------	--------------

Table1 :Summary of survey

References:

1. A.I.Awad, N.A.El-Hefnawy and H.M.Abdel_kader,” Dynamic Multi-objective task scheduling in Cloud Computing based on Modified particle swarm optimization”, *Advances in Computer Science: an International Journal*, Vol. 4, Issue 5, No.17 , September 2015.
2. Atul Vikas Lakraa, Dharmendra Kumar Yadav ,” Multi-Objective Tasks Scheduling Algorithm for Cloud Computing Throughput Optimization”, *International Conference on Intelligent Computing, Communication & Convergence*, *Procedia Computer Science* 48 (2015) 107 – 113
3. Majid Habibi, Nima Jafari Navimipour,” Multi-Objective Task Scheduling in Cloud Computing Using an Imperialist Competitive Algorithm”, *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 5, 2016.
4. R K Jena,” Multi objective Task Scheduling in Cloud Environment Using Nested PSO Framework”, *Procedia Computer Science* 57 (2015) 1219 – 1227.
5. Shikha Chaudhary ,Saroj Hiranwal ,C. P. Gupta,” Review on Multiobjective Task Scheduling in Cloud Computing using Nature Inspired Algorithms”, *International Journal of Emerging Research in Management &Technology*, ISSN: 2278-9359 (Volume-6, Issue-8)
6. Shubhashree S. R.,” Multi-Target Tasks Scheduling Algorithm for Cloud-environment Throughput Optimization”,*International Journal of innovative research and development*,Volume 5 Issue 8,July 2016(ISSN:2278-0211)
7. Sourabh Budhiraja, Dr. Dheerendra Singh, “An Efficient Approach for Task Scheduling Based on Multi-Objective Genetic Algorithm inCloud Computing Environment”, *JCSC VOLUME 5 • NUMBER 2 JULY-SEPT 2014 PP. 110-115 ISSN-0973-7391*.
8. Vanita Dandhwani, Dr.Vipul Vekariya,” Multi-Objective Task Scheduling using K-mean Algorithm in Cloud Computing”, *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 4, Issue 11, November 2016.
9. VG.Ravindhren and Dr. S. Ravimaran ,” Responsive Multi-objective Load Balancing Transformation Using Particle Swarm Optimization in Cloud Environment”, *Journal of Advances in chemistry*, Volume 12 Number 15, ISSN 2321-807X.
10. Xhafa F, Abraham A., “Computational models and heuristic methods for Grid scheduling problems”, *Futur Gener Comput Syst*, Volume 26, Issue 4, pp. 608–621, April, 2010.
11. Y.Chawla, M.Bhonsle, “A Study on Scheduling Methods in Cloud Computing”, *International Journal of Emerging Trends & Technology in Computer Science*, Volume 1, Issue 3, pp. 12-17, September –October, 2012.

Novel Evaluation of Protein Convertase Substilisin/Kexin Type 9 (PCSK9) Gene by Motif Detection Technique

Raiha Tallat
CS.Department
(CIIT) Sahiwal
Sahiwal, Pakistan
raiha.tallat@live.com

M.Shoaib
CS. Department
(CIIT) Sahiwal
Sahiwal, Pakistan
mshoaib@ciitsahiwal.edu.pk

Javed Ferzund
CS.Department
(CIIT) Sahiwal
Sahiwal, Pakistan
jferzund@ciitsahiwal.edu.pk

Ahmad Nawaz Zaheer
MS.Department
(GCUF) Sahiwal
Punjab, Pakistan
anzaheer@hotmail.com

Sana Yaseen
CS. Department
(CIIT) Sahiwal
Sahiwal, Pakistan
sanayaseen42@yahoo.com

Umar Draz
CS.Department
(CIIT) Sahiwal
Sahiwal, Pakistan
sheikhumar520@gmail.com

Abstract— with the emerging of new development in solving the issues related to big data and its important in computer science cannot be undermined. In this research paper we described briefly the motif occurrence and uniqueness for a significant PCSK9 gene responsible for the generation of protein which causes lower cholesterol levels. Motif occurrence is very common in protein sequences and their detection provides a very important role in evaluating the function of proteins. Currently many motif databases are present which help us comparing the specific motifs with currently available motifs, results in the paper are compared with 11 available databases associated with the TOMTOM tool. Results are calculated using three different clusters of PCSK9 protein sequences. 48 different species are the members of the clusters used in the analysis done by the help of de novo algorithm used by the MEME suite. The approach can be considered as the novel example of the renowned problem of motif detection in large graphs using big data analytic techniques.

Keywords-motifs; PCSK9; MEME; transcription; bioinformatics; Genomics

I. INTRODUCTION

In the era where data has raised to a huge amount, The emerging size of data is opening up the new possibilities for the data scientists in every field whether it's Mechanics, Bioinformatics, Genomics, Media, Business, Computer sciences, Electronics, Health sciences, Telecommunication etc [1]. As it keeps growing it needs to be analyzed so that the hidden information underneath can be fetched and useful decisions can be made on its base. Literature shows that data scientists have developed many algorithms regarding data visualization and analysis to make it more presentable and interpretable for the sake of information retrieval [2]. Data analysis is overlapped with visual analytics because visual

analysis has its significant place in the world of Big Data [3]. One of the appropriate techniques for representing data is in the form of graphs. The nodes in a graph can represent the prominent entities depending upon the type of data being processed. According to literature visual graph analysis has gained the attention of researchers in order to process various data formats. One of the paradigms of Big Data is that few algorithms provide efficient results but with limited scalability of data. In case of massive data sets for example protein sequences and DNA sequences, algorithms with more accuracy are required in order to obtain results. Biologists have been working on massive datasets using the techniques of Big Data analytics [4]. In large dynamic graphs often repetitions of patterns are occurring which show that some specific path is being followed repeatedly. This sub graph or repetitive path is known as the "Network Motif" [5]. In case of protein sequences these repetitions are found abundantly thus motif occurrence becomes a factor of similarity. Repetitive patterns in proteins show the function similarity. This similarity holds a special importance because proteins are mutating with the passage of time which causes various changes in different species. These mutations can be compared thus creating connectivity which identifies the similarity between protein structures [6]. Random projections are used for the discovery of motifs with the help of an algorithm named PROJECTION [7]. Other than PROJECTION an algorithm named de novo motif finder is also currently the shining star in network motif detection thus showing the best motifs found in a sequence and their comparison with other motif databases. Among different gene sequences the PCSK9 gene is very important because this gene is responsible for the generation of protein which lowers the level of cholesterol in the blood stream thus resulting into a decreased rate of cardiac

diseases [8]. The PCSK9 helps breaking down the low-density lipoproteins receptors which are carriers of cholesterol in the bloodstream.

II. RELATED WORK

Motif detection is performed using a number of different algorithms using DREME, MAST, GYM, DMINDA² etc. The DREME algorithm is used for the discovery of motifs based on transcription factor (TF). It allows large data sets to be analyzed and obtains various binding motifs in sequences. Another program used for motif detection is GYM. GYM is known for the Helix-turn-Helix motif detection. Helix-turn-Helix motifs are among the widely studied motif structures as per literature [9], moreover GYM also provides comprehensive information on the protein sequences. Some approaches of artificial intelligence have also been used for the discovery of motifs in gene sequences, one such example is the MAST algorithm, which conveniently uses output from MEME [10] for searching databases such as SWISS-PROT and Genprot. MAST projects some statistical measures that permit a rigorous evaluation of the significance of database searches with individual motifs or groups of motifs [11]. Another facility for the detection of motifs is DMINDA, now known as DMINDA², it's an integrated web browser used for the discovery of motifs in the given sequences. The interface of DMINDA² also provides the location of a given motif in a sequence [12]. It provides a suite of *cis*-regulatory motif analysis functions on DNA sequences. DMINDA² follows four steps for the DNA sequences analysis:

III. PROPOSED METHODOLOGY

In this section the methodology for the sequence analysis is mentioned. Three clusters of sequence s having 48 different species are downloaded from the UniProt database. All the sequences were in FASTA format and unaligned. The sequences were than uploaded to t he MEME for motif discovery with specific parameters. Minimum width for motif detection was 6 and maximum width w as 50. The sequence of the analysis is represented by the given diagram. Starting from the outer circle and moving towards the target. The goal is to find a motif unique in nature. Identification of a unique motif means that motif can be studied for biologists and it has some specific functionality which is not present in the existing databases.

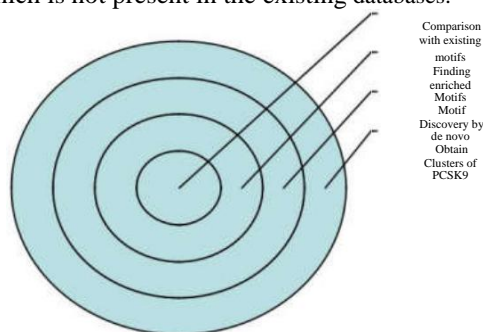


Figure 1: Analysis Methodology

IV. RESULTS AND CONCLUSIONS

This section represents results obtained from 3 different clusters of PCSK9 gene sequence. MEME suite was used for the discovery, locating enrichment and comparison of motifs.

A. DeNovo motif detection by MEME

The figure shown below represents the de novo motif discovery using the MM algorithm and d comparison of motifs using the TOMTOM technique across existing motif databases.

Database	ID	Alt. ID	Preview	Matches
meme	NAACQR	MEME-3		0
meme	NFGDDVDLYAP	MEME-1		0
meme	PEDITGGASDAQDQPV	MEME-2		0

Figure 2: DeNovo motif detection using MM Algorithm

Figure 3 represents the results calculated by the MAST algorithm used in the MEME suite.



Figure 3: Initial and ending point of the identified motif using MAST Algorithm



Figure 4: All related motifs in network PCSK9 gene

The figure given below shows the motif with minimum E-value of 1.1e+001 in the PCSK9 protein sequence. The variation in the size of the alphabets representing amino acids in the motif shows conservation.

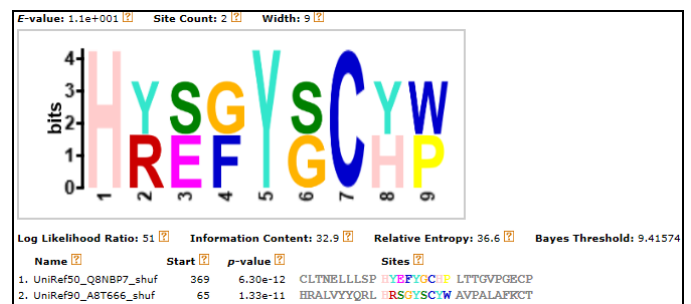


Figure 2: Motif with minimum e-value in PCSK9 sequence

B. The Bayes Threshold:

The bayes threshold of the considered motif is calculated using the Bayesian optimum classifier. Our results showed that the naïve classifier performed efficiently for detecting motifs with the sequence having range of 400 residues. It also performed well with the controlled sequences which were the shuffle version of the original sequences in order to observe motif conservation.

CONCLUSION

The approach can be considered as the novel example of the renowned problem of motif detection in large graphs using big data analytic techniques. After discovering motifs in PCSK9 sequences the next step was determining the uniqueness and enrichment of the motifs. This paper describes the motif occurrence and uniqueness for a significant PCSK9 gene responsible for the generation of protein which causes lower cholesterol levels. Results were concluded with the fact that PCSK9 gene obtained from 48 different species had the smallest motif with length 49 which represents a significant amount of similarity. Further comparison with existing TOMTOM databases showed that only 1 motif was determined unique in nature thus representing the mutation caused by PCSK9. Few motifs had similarity with other motifs but similarity index was limited as compared to other motifs. Detected motifs can be further used for finding the binding pockets of the proteins thus determining its functionality.

REFERENCES

- [1] Wernicke, S., & Rasche, F. (2006). FANMOD: a tool for fast network motif detection. *Bioinformatics*, 22(9), 1152-1153.
- [2] Moses, A. M., Chiang, D. Y., & Eisen, M. B. (2003). Phylogenetic motif detection by expectation-maximization on evolutionary mixtures. In *Biocomputing2004* (pp. 324-335)
- [3] Vahdatpour, A., Amini, N., & Sarrafzadeh, M. (2009, July). Toward Unsupervised Activity Discovery Using Multi-Dimensional Motif Detection in Time Series. In *IJCAI* (Vol. 9, pp. 1261-1266).
- [4] Xing, E. P., Wu, W., Jordan, M. I., & Karp, R. M. (2004). LOGOS: a modular Bayesian model for de novo motif detection. *Journal of Bioinformatics and Computational Biology*, 2(01), 127-154..
- [5] Wong, E., Baur, B., Quader, S., & Huang, C. H. (2011). Biological network motif detection: principles and practice. *Briefings in bioinformatics*, 13(2), 202-215.
- [6] Dodd, I. B., & Egan, J. B. (1990). Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic acids research*, 18(17), 5019-5026.
- [7] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594), 824-827.
- [8] Dodd, I. B., & Egan, J. B. (1990). Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic acids research*, 18(17), 5019-5026.
- [9] Harrison, S. C., & Aggarwal, A. K. (1990). DNA recognition by proteins with the helix-turn-helix motif. *Annual review of biochemistry*, 59(1), 933-969.
- [10] Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... & Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, 37(suppl_2), W202-W208.
- [11] Bailey, T. L., Baker, M. E., & Elkan, C. P. (1997). An artificial intelligence approach to motif discovery in protein sequences: application to steroid dehydrogenases. *The Journal of steroid biochemistry and molecular biology*, 62(1), 29-44.
- [12] Ma, Q., Zhang, H., Mao, X., Zhou, C., Liu, B., Chen, X., & Xu, Y. (2014). DMINDA: an integrated web server for DNA motif identification and analyses. *Nucleic acids research*, 42(W1), W12-W19.

A Review on Detection and Counter Measures of Wormhole Attack in Wireless Sensor Network

Rabia Arshad, Saba Zia

Abstract- Sensor nodes are organized to form wireless sensor network to be deployed in hostile environments. Sensor nodes communicate each other routing protocols. Information from source node to destination node is sent via intermediate nodes. Security is a major issue in WSN in present days as WSN is vulnerable to attacks that can cause damage to the functionality of the system. In this survey paper an attempt has been made to analyze security threats to WSN at network layer. Network layers is affected by many attacks e.g. Black Hole Attack, Grey Hole Attack, Wormhole Attack, out of which Wormhole attack is the most devastating where attacker agents make a link between two points with low latency. This paper focuses on some researches in detecting and preventing the wormhole attack in network layer.

Keywords- *Mobile Adhoc Network, Security, Wireless Sensor Network, Wormhole Attacks*

I. INTRODUCTION

WSN [1-5] is composed of numerous sensor nodes that are capable of monitoring environmental conditions. Sensor nodes are responsible for transmitting the information in the network. Transmitting data in the WSN is a critical task because sensor nodes are restricted devices. Due to this reason, sensor network is susceptible to many attacks. WSNs have some special features that distinguish them from other networks. These characteristics are given as follows [18]:

- Limited resources
- Minimum battery life span
- Self-configuration
- Random changes in topology of network
- Centralized approach of network control

Security is one of the important challenges in designing the WSN. Data is vulnerable to attacks of many kinds therefore security measurements should be taken while designing the WSN. Many security attacks can affect the performance of WSN e.g. Black hole [19], Grey hole [20], Wormhole attack. Wormhole attack is the most dangerous attack to WSN. From this point of view, this paper briefly describes the techniques for detection and prevention of wormhole attack. Rest of the paper is described as follows:

- Section 2 describes about the challenges of WSN.
- In section 3 various attacks on WSN are summarized.
- Section 4 covers the background of Wormhole attack
- Section 5 describes the different types of Wormhole attack
- In section 6, different modes of wormhole attack are discussed
- Section 7 listed some counter measures to wormhole attack

II. CHALLENGES OF WSN

According to different application requirements the following design objectives of sensor nodes are considered [16, 17]:

A. *Low Cost Node*

Sensor nodes are deployed usually in a harsh environment in a large quantity. Also the sensor nodes are not reusable. Therefore, reducing the cost of the sensor nodes is an important step in network design.

B. *Low Cost Node*

Reducing the size of the sensor node reduces the power consumption as well as also the cost of sensor nodes. Reducing the node size is very useful in the node deployment in hostile environment.

C. *Low Energy Consumption*

Sensor nodes consume power in performing their function. Power in the sensor nodes is provided by the batteries. In some situations it is impossible to charge their batteries. Therefore reducing the power usage of sensor nodes is a crucial factor. In this way, the network lifetime can be prolonged.

D. *Scalability*

Routing protocols for the sensor networks must be scalable to different network sizes as the sensor network consist of thousands of sensor nodes.

E. *Reliability*

Protocols for sensor networks must include error detection and correction techniques. By these techniques, a protocol ensures the reliability of data delivery over some noisy link.

F. *Adaptability*

In sensor networks, any fault can occur in the network due to which a node may fail. A new node may be added in the network at some later stage. It is also possible that a node may move to some new place in the network (in the mobile network). These situations result in variations in the network topology. Therefore, the network protocols should be adaptive to such changes in the network.

G. *Channel Utilization*

The network protocols should efficiently use the bandwidth to improve utilization of the channel as the sensor networks have limited bandwidth resources.

H. *Fault Tolerance*

WSN is mostly deployed in the harsh or hostile environment. Nodes might get failed due to the harsh environmental conditions. Therefore, sensor nodes must have the characteristics of fault tolerant.

I. *Security*

Information in the network must be secured and prevented from malicious attacks. Thus, effective security techniques must be introduced in the sensor network to avoid these kinds of attacks.

III. ATTACKS ON WSNS

WSN are susceptible to a variety of attacks because of multi-hop transmission. As nodes are deployed in hostile environment therefore WSNs have some additional vulnerability to attacks. Table I summarizes the possible attacks on different layers of WSN and their possible solutions.

TABLE I. ATTACKS ON DIFFERENT LAYERS ON WSN AND THEIR SOLUTIONS

Layer	Attacks	Security Solutions
Physical	Tampering	Tamper Proofing Hiding Encryption
Data Link	Jamming Collision	Error correction method Spread Spectrum Method
Network	Sybil Sinkhole Wormhole	Authentication Authorization
Transport	Packet Injection	Packet Authentication
Application	Aggregation Based Attacks	Cryptographic Approach

IV. NETWORK LAYER ATTACKS AND THEIR EFFECTS

Layered architecture of WSN make it more vulnerable to security attacks. Various attacks and their defensive techniques have been proposed in WSN. Attacks on network layer [6] are given in Table II.

TABLE II. NETWORK LAYER ATTACKS IN WSN AND THEIR EFFECTS

Attack	Description	Effects
Wormhole	a. Require 2 or more adversaries b. These adversaries have better resources of communication between them [7].	<ul style="list-style-type: none"> • Network topology changes. • Packets are destructed. • False information for routing.
Sybil	A malicious node represents different identities and attracts the traffic [8].	<ul style="list-style-type: none"> • Disruption of WSN • Can be a source for other attacks
Black Hole	A malicious node behaves like destination node and does not forward the packet.	<ul style="list-style-type: none"> • Throughput is decreased [7] • Disruption of WSN
Sink Hole	More complex than Black Hole [7]	<ul style="list-style-type: none"> • Attracts all the traffic • Other attacks are also triggered. • Base station position is affected.
Selective Forwarding	A malicious node selectively drops the packet and does not forward. It acts like Black Hole [9].	<ul style="list-style-type: none"> • Message contents are modified. • Packet dropping • Resources can be exhaustive.
False Routing	Here the attackers creates the loops in the network by routing packets to false sink node [10].	<ul style="list-style-type: none"> • False messages • Resource exhaustion

V. BACKGROUND AND SIGNIFICANCE OF WORMHOLE ATTACK

WSN is vulnerable to attacks of different types due to the scarcity of resources. Wormhole attack is a severe attack on network layer of WSN where two or more attacking agents are connected by high speed wormhole off-channel link [12]. Wormhole attack has two different mode of attacking i.e. ‘Hidden’ and ‘Exposed’ mode. In exposed mode of attack, identity of attacker is attached in the packet header while tunneling and replaying packets [11].

In wormhole attack, any two attackers form a tunnel to transfer data and replays this data in the network. This tunnel is referred to as wormhole. Wormhole attack effects the WSN tremendously. Routing techniques may be disrupted when routing messages are tunneled. Figure. 1 represent a scenario of wormhole attack. Packets received at node A are replayed via node B and vice versa.

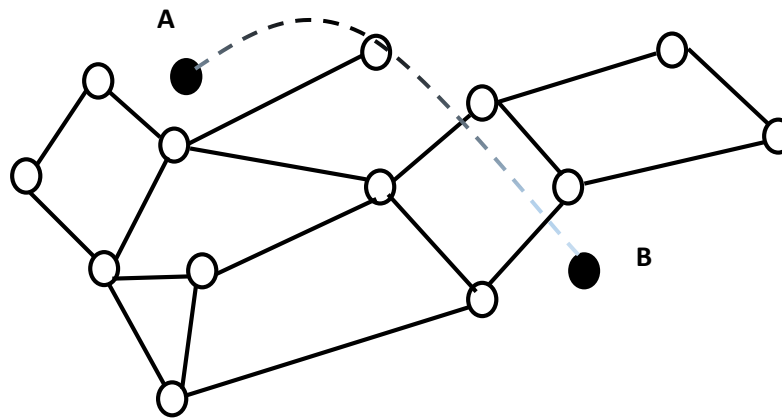


Figure 1. Wormhole Attack

VI. CLASSIFICATION OF WORMHOLE ATTACK

Wormhole attack is very difficult to detect. It is one of the Denial of Service attacks. Wormhole attack might be launched by any number of nodes. It is required to categorize the wormhole attacks for detection and prevention of wormhole attacks. Wormholes are divided into three categories [13] i.e. *Open*, *Closed* and *Half Open*. Table III describes which nodes are visible or invisible in three types of classifications. Figure 2, 3 and 4 describe the three types of wormhole attack respectively where S= source, D= destination, M= Malicious Node. Classification of wormhole attack is based on following:

- i. Attackers are invisible/ visible
- ii. Data forwarding mechanism of wormhole nodes
- iii. Ability to hide and show the identities of nodes

TABLE III. NODES DESCRIPTION IN DIFFERENT WORMHOLE ATTACKS

	Open	Half Open	Closed
Source Node	Visible	Visible	Visible
Destination Node	Visible	Visible	Visible

Malicious Node1	Visible	Visible	Invisible
Malicious Node2	Visible	Invisible	Invisible

A. Open Wormhole

Source node, destination node and malicious nodes are visible in open wormhole. Nodes X and Y are kept hidden on traversing path. In this mode, packet header also contains the attackers. All sensor nodes in the network are aware of the presence of malicious nodes and would represent the malicious nodes as direct neighbors.

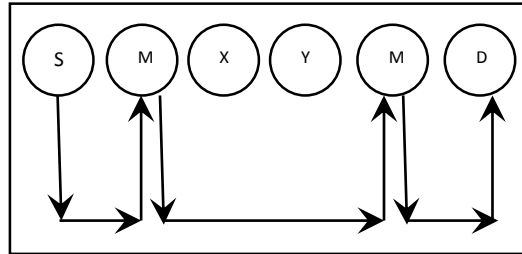


Figure 2. Open Wormhole

B. Closed Wormhole

In this mode, source node and destination node behaves one hop away from each other and this leads to creation of fake neighbors.

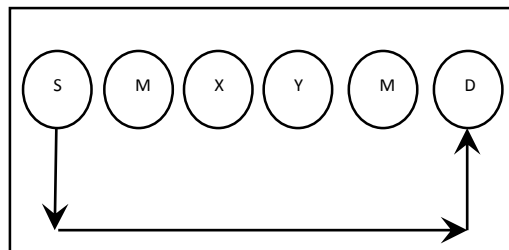


Figure 3. Closed Wormhole

C. Half Open Wormhole

The first end of malicious node is visible near source node and second end is hidden in this mode of wormhole attack. Contents of the data packet are not modified by attackers. Packets are rebroadcast in this type of attack because the attackers tunnel the packets from one malicious end to other.

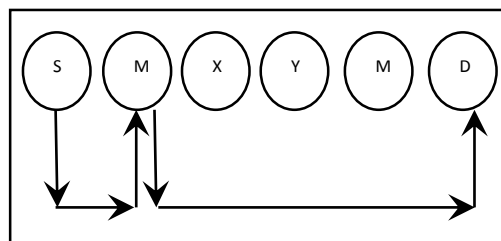


Figure 4. Half Open Wormhole

VII. MODES OF WORMHOLE ATTACKS

Wormhole is categorized into the following types based on how many nodes are involved in establishment of wormhole.

A. Wormhole Using Protocol Distortion

This type of wormhole does not affect the network functioning too much so it is a harmless type. It is also known as “Rushing Attack”. Only a single node is a malicious node that distorts the routing protocol.

B. Wormhole using Packet Relay

In this type of attack, malicious nodes replay data in between two nodes at far distance from each other. This would lead to creation of fake nodes. It is also “Replay Based Attack”.

C. Wormhole using Out-of-Band Channel

In this type of attack, there is only one malicious node capable of high transmission. It attracts the data packets to traverse the route that passes from the malicious node.

D. Wormhole using Packet Encapsulation

In this type of attack, there are many nodes presents between any two malicious nodes. Data is encapsulated in between malicious nodes.

Table IV summarizes different modes of wormhole attack and number of adversary nodes is given [13]:

TABLE IV. SUMMARY OF WORMHOLE ATTACK MODES

Mode	Minimum Adversary Nodes
Protocol Distortion	1
Packet Relay	1
Out of Band Channel	2
Packet Encapsulation	2

VIII. COUNTER MEASURES OF WORMHOLE ATTACK

In this section, some important wormhole detection methods are discussed. Table V summarizes a description of wormhole detection methods.

A. Statistical Analysis Method

Song et al. proposed a mechanism for detection of wormhole based on statistical analysis of multipath routing. This method is useful for multipath and on-demand routing protocols in intrusion detection systems [14].

B. Graph Theory Method

Graph theory method was proposed by Lazos and Poovendran [11] for detection of wormholes. In this method, nodes that are used have information about location and thus are called location-aware guard nodes (LAGNs). LAGNs use a key called “local key” between neighbors that are one hop apart. The message that is encrypted with local is not possible to decrypt. Therefore hashed messages are used in formation of local key to detect wormholes. Node will detect a number of inconsistencies in the message if there is a wormhole otherwise it will unable to response.

C. Hop Counting Method

Hidden as well as exposed attacks can be detected using this method. In this method, DELPHI [14] (Delay per Hop Indicator) protocol is used. Length of each route and time delay for each route is calculated for identification of wormhole. Therefore, the route having wormhole faces a greater time delay as compared to other routes.

D. Visualization Based Method

In this method [11], each node computes the distance to its neighboring nodes by using received signal strength. According to these calculations, networks topology is calculated by the base station. Network topology is more or less flat in case there are no wormholes. In wormholes are present then in visualization it can be seen there is a string at different ends of network that pulls it. In this method, each node has to send its list of neighbors to the base station.

E. Hardware Based Method

A method based on directional antennas [15] was proposed by Hu and Evans. It was based on assumption that in absence of any wormhole, if node A sends data in particular direction than it was received at its neighbor from opposite direction. It is mandatory that every node must have its directional antennas.

F. Trust Based Method

Jain and Jain proposed another important method [14] for identification and isolation of malicious nodes that create wormhole in the network. In this method, trust values are calculated according to the sincerity of the nodes in the neighbors for execution of routing protocol. The trust value is used to effect the routing decisions which guide the nodes not to communicate through wormholes. Packet dropping is reduced to 14% by using the trust based methods. Using trust based mechanisms, throughput is also increased up to 8-9%.

TABLE V. SUMMARY OF METHODS FOR WORMHOLE DETECTION

Methods	Description
Statistical Analysis Method	Efficient for on demand protocols, easy integration and effective for multipath routing
Graph Theory Method	Nodes are equipped with GPS receivers
Hop Counting Method	Low overhead, high efficiency and fast performance
Visualization Based Method	Mobility is not considered in this method. It is best for dense networks
Hardware Based Method	Not applicable to the networks other than having directional antennas. Very efficient for the networks having antennas
Trust Based Method	Locate dependable routes in the network in an effective way

IX. CONCLUSION

In this paper, attacks to WSN are studied briefly. Wormhole attack is one of the important attacks on WSN. Background of wormhole attack, its cause's and methods of prevention i.e. graph theory method, trust based method, and hop-counting methods etc. are discussed in this paper. Graph theory method requires GPS receivers it is location based method. Hardware method is the efficient method for large networks having antennas. Hop-counting method have high efficiency over other methods.

REFERENCES

- [1] Neha Rathi, Jyoti Saraswat and Partha Pratim Bhattacharya, "A REVIEW ON ROUTING PROTOCOLS FOR APPLICATION IN WIRELESS SENSOR NETWORKS". *International Journal of Distributed and Parallel Systems*, Vol. 3, No. 5, September 2012
- [2] Kemal Akkaya, Mohamed Younis, "A Survey on Routing Protocols for Wireless Sensor Networks". *Ad Hoc Networks* (2005)
- [3] Abdalaleem Ali Almazroi, MA Ngadi, "A Review on Wireless Sensor Networks Routing Protocol: Challenges in Multipath Techniques" *Journal of Theoretical and Applied Information Technology*, Vol. 59, No. 2, January 2014
- [4] Rajashree, V. Biradar, V. C. Patil, Dr. S. R. Sawant, Dr. R. R. Mudholkar, "Classification and Comparison of Routing Protocols in Wireless Sensor Networks". *Special Issue on Ubiquitous Computing Security Systems*.
- [5] Laiali Almazaydeh, Eman Abdelfattah, Manal Al- Bzoor, and Amer Al- Rahayfeh, "Performance Evaluation of Routing Protocols in Wireless Sensor Networks". *International Journal of Computer Science and Information Technology*, Volume 2, Number 2, April 2010
- [6] Nityananda Sarma , Sangram Panigrahi, Prabhudutta Mohanty and Siddhartha Sankar Satapathy, "Security Issues in Wireless Sensor Network Data Gathering Protocols: A Survey". *Journal of Theoretical and Applied Information Technology*, 2005
- [7] Kia Xiang, Shyaam Sundhar Rajamadam, Srinivasan, Manny Rivera, Jiang Li, Xiuzhen Cheng, "Attacks and Countermeasures in Sensor Networks: A Survey", *Springer*, 2005
- [8] Sushma, Deepak Nandal, Vikas Nandal, "Security Threats in Wireless Sensor Networks". *International Journal of Computer Science & Management Studies*, Vol. 11, Issue 01, May 2011
- [9] Syed Ashiqur Rahman, Md. Safiqul Islam, "Anomaly Intrusion Detection System in Wireless Sensor Networks: Security Threats and Existing Approaches". *International Journal of Advanced Science and Technology*, Vol. 36, November 2011
- [10] Ali Modir Khazeni, Norafida Ithnin, Mohammad Javad Abbasi, "Secure Hierarchical Routing Protocols in Wireless Sensor Networks: Security Survey Analysis", *International Journal of Computer Communications and Networks*, Volume 2, Issue 1, February 2012
- [11] Majid Meghdadi, Suat Ozdemir and Inan Guler , "A Survey of Wormhole Based Attacks and Their Countermeasures in Wireless Sensor Networks", *IETE TECHNICAL REVIEW*, Vol. 28, ISSUE 2, April 2011
- [12] Khin Sandar Win, "Analysis of Detecting Wormhole Attack in Wireless Networks". *World Academy of Science, Engineering and Technology*, 2008
- [13] Marianne Azer, Sherif El-Kassas, Magdy El-Soudani, "A Full Image of the Wormhole Attacks Towards Introducing Complex Wormhole Attacks in Wireless Ad Hoc Networks", *International Journal of Computer Science and Information Security*, Vol. 1, May 2009
- [14] Preeti Nagrath, Bhawna Gupta, "Wormhole Attacks in Wireless Adhoc Networks and their Counter Measurements: A survey". *IEEE*, 2011
- [15] Zhibin Zhao, Bo Wei, Xiaomei Dong, Lan Yao, Fuxiang Gao, "Detecting Wormhole Attacks in Wireless Sensor Networks with Statistical Analysis". *IEEE International Conference on Information Engineering*, 2010
- [16] Jamal Al-Karaki and Ahmed E. Kamal, "Routing Techniques in Wireless Sensor Networks: A Survey". *IEEE Communications Magazine*, Vol. 11, No. 6, December 2004
- [17] Kemal Akkaya and Mohamed Younis. "A Survey on Routing Protocols for Wireless Sensor Networks", *Ad hoc Networks*, Vol. 3, No. 3, May 2005
- [18] Shio Kumar Singh, M P Singh, and D K Singh, "Routing Protocols in Wireless Sensor Networks – A Survey". *International Journal of Computer Science & Engineering Survey*, Vol. 1, No.2, November 2010
- [19] Arshdeep kaur, Mandeep kaur, "A Survey Black Hole Attack in Manet". *International Journal of Science, Engineering and Technology Research*, Volume 4, Issue 5, May 2015
- [20] Mitali Khandelwall, Sachin Upadhyay, "Detecting and Preventing Black hole and Grey Hole Attacks for Trust Management in Wireless Sensor Networks: A Survey". *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 5, Issue 5, May 2016

Performance Analysis of spot diffusion technique for indoor optical communication system employing LDPC

Sabira Khanam Shorna

Department of Computer Science & Engineering,
Primeasia University,
Dhaka, Bangladesh
Email: sabira001840@gmail.com

Koushik Chandra Howlader

Dept. of CSTE,
Noakhali Science and Technology University,
Noakhali, Bangladesh
Email: koushiksoul@gmail.com

Arup Mazumder

Department of Electronics and Telecommunication Engineering,
Southeast University,
Dhaka, Bangladesh
Email: arup.mazumder27@gmail.com

Abstract— Indoor optical wireless communication systems offer an attractive substitution for realizing next generation wireless local area network. In the context of indoor optical-wireless communication (OWC) system the spot-diffusing technique provides improved performance compared to conventional diffuse system. In this work, the performance of an OW spot-diffusing communication system using Neuro-Fuzzy (NF) adaptive multi-beam transmitter configuration has been proposed. This research work focus on the performance of the indoor optical wireless systems. Where BER of the existing modulation technique evaluated based on the Spot Diffusion Adaptive ANFIS algorithm and compared the existing result and analysis. A Linear error correcting code low-density parity-check (LDPC) scheme is introduced in the algorithm to observe the change in the performance of the system. It has been analyzed that involving LDPC in Spot Diffusion Adaptive ANFIS algorithm provides approximately 37.5% improvement of BER result against receiver position, where BER reduce chronologically. From the bit error rate analysis of these schemes it has seen that the system model with LDPC performs better than other existing techniques.

Keyword: - Indoor optical wireless communication (OWC), Neuro-Fuzzy (NF) adaptive multi-beam transmitter, Spot Diffusion Adaptive ANFIS algorithm, Linear error correcting code low-density parity-check (LDPC).

I. INTRODUCTION

This work is focusing on the performance analysis for the BER of indoor optical wireless communication system using LDPC. Some of advantages of OWC are low cost, base-band circuit design, High data rates (Gbps), less multi access interference, no need to pay for spectrum license etc. and limitations are it can't pass through wall, sensitive to blocking; limited Transmit Power etc. These research works have their own features, possibilities and limitation which are described later in the paper. However wireless systems have some working and performance limitation. Some key points are discussed below:
Optical fiber communication transmits information through optical fibers is largely replaced by radio transmitter systems for long-haul optical data transmission. Not only telephony but also Internet traffic, long high-speed local area networks (LANs), cable TV (CATV) has been used such systems. OWC is defined as the use of optical frequencies to carry the electrical signals. Unguided visible, infrared (IR), or ultraviolet (UV) light to carry a signal.

For indoor communication though infrared provides significant advantages as a medium but it also has some drawbacks. Several aspects impair the performance of indoor IR transmission systems. Because of such hinders design and implementation is not so easy using infrared. So, the optical wireless communication is used to overcome such indoor environment.

Following table 1.1 gives a comparison between optical wireless and radio frequency systems.

Criteria	Optical wireless systems	Radio systems
Bandwidth	Unregulated large	Limited
Passes Through wall	No	Yes
Cost	Low	High
Speed	High	Low
Fading	Free from fading	Multipath fading.
Security	Security and freedom from spectrum regulation and licensing.	Low security.

Table 1.1: Comparison Between Optical Wireless and Radio System in Indoor Wireless Communication.

This can be used for different ranges:

- Short range (cm – m): Chip-to-Chip Interface,
- Medium range (m – 10 m): Wireless Optical LAN,
- Long range (km): Free-Space Optical Communications.

However, In Optical wireless LAN where it can be used to illuminate the room while serving as a medium for data transfer and also for transferring data at high speed for long distance. For free space communication, it also offers high speed long distance data transfer using satellite.

1. Eye safety consideration puts limit on the amount of optical power that should be emanated by the transmitter, thus limiting the coverage of an optical wireless system.
2. In indoor optical wireless systems, the leading source of noise is ambient light, which is typically a combination of fluorescent light, sunlight, and incandescent light. Ambient light provokes shot noise due to the random nature of the photo-detection process.
3. A multipath phenomenon occurs when the transmitted signal follows different paths on its way to the receiver due to its reflection by walls, ceilings and other objects. Channel dispersion associated with multipath propagation is another major issue in indoor optical wireless systems. Multipath phenomena can cause inter-symbol-interference (ISI).

During implementing the performance of optical wireless system in indoor context, some problems were noted. Especially evaluation of BER is major factor for such systems.

another technique named LDPC have used for long-haul data transmission using the principal of forward-error

correction in the optical wireless communication. Another work where using Neuro fuzzy technique analysis the BER according to the mobility of the transmitter and receiver. But here BER have not investigated for the respected distance between transmitter and receiver.

Thus, investigating the BER for the allowing distance between transmitter and receiver in indoor context of optical wireless communication is the major issue of this proposed work, where reason of using the LDPC technique is to improve the BER performance then previous work and study.

- Previous work of indoor optical wireless communication has analyzed mainly the concern of BER. All the information that are related to this proposed research are investigate and simulate in Matlab platform.
- If the simulation could have been done in real environment, more appropriate result might have been measured.
- Here considered the upward light illuminating measurement. The measurement in a room for the side wall reflection are not listed.
- Five parameters are considered by the proposed research, they are SNIR (Signal –to-Noise-Interference), Delay Spread, BER with and without LDPC. There are more parameters that could have been compromised.

The objectives of the study are:

- to review on existing indoor optical wireless communication system improvement of BER
- to propose a method like LDPC in this area that look up and analysis the BER for the system.
- to compare the performance of proposed method with existing research

All of this aim would cover the enough pretty analysis of BER for high-speed indoor OW applications.

In chapter 2, a number of different proposed works are discussed based on research papers. Proposed system model has been discussed in chapter 3. Numerical analysis and comparison with other methods have been established in chapter 4.

II. LITERATURE REVIEW

In this chapter, there are brief knowledge about the OWC and BER. Many studies were introduced using this method for different wireless communication from time to time. Here many research are discussed and show comparison among them.

Today's world research heeds their interest mostly in the context of wireless communication. The lofty maintenance and configuring cost makes the wireless system financial and flexible alternative to wired system. People are aware of a proliferation of many researches and developments that has been done in this area specifically in Optical wireless communication. Further generation of wireless communication systems are already relay on Optical Wireless System technologies. Conventionally OWC technology depends on high power solid-state lasers or diode lasers for medium to long-range applications. Latterly we have seen the remarkable advances in semiconductor sources such as light emitting diodes (LEDs) in visible light and ultraviolet wavelengths, multi-array light sources and detectors, tracking and steering. With the benefits of low power and cost for short/medium range wireless communication applications such advances provide huge latent. The optical wireless Communication could provide a cost effective, flexible, secure and ultra-high-speed solution to the materializing challenges facing the system and service providers over RF (Radio frequency).

Optical wireless communications have becoming an effective alternative medium to optical fiber, and radio frequency (RF) communications and it optimistically removes gradually all of the possible hinders that are challenging for previous communication technologies,

because of its high bandwidth, low cost, ease of implementation, license free spectrum freedom from interference and many more. For some wireless applications such as 3-D face-to-face communication and super Hi-Vision/Ultra High Definition TV data (more than 4Gbit/s) [1] due to the explosive bandwidth which can be envisioned through optical wireless communication. Data throughput and transmission link based on optical wireless these are the great concern for number of applications [2] [3].

Most of the time the transmitted data it is not fully secure or error free. Some statistical fluctuations are related to noise influences (e.g. laser noise, amplifier noise, shot noise, or excess noise of a receiver) cause a small fraction of the transmitted bits to be defective. Typically, the bit error rate (i.e., the fraction of incorrectly transmitted bits) is strongly dependent on the transmitted power, and the latter must be high enough to keep the bit error rate below a certain acceptable level (e.g. 10^{-12} for Earth-based telecommunication systems, or 10^{-6} for satellite control). Nearly all of the remaining bit errors can be detected using some kind of checksums and corrected.

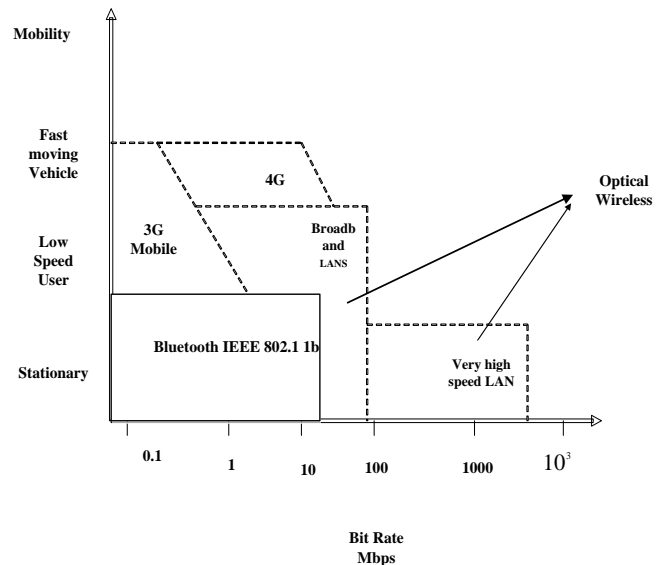


Fig 2.1: Simple Scenario of BER in Optical Wireless communication

The error correction scheme can use some level of redundancy in the transmitted data or involve retransmitting corrupted data packets. Additional detrimental influences such as fiber losses or various types of dispersion in a longer link, or background light in free-space transmission, can often be compensated by somewhat increasing the transmitted power. BER is the total number of erroneous bits compared to the total number of transmitted bits.

The increase of optical power required to maintain a given bit error rate is called a *power penalty*, or more specifically e.g. a *dispersion penalty* if dispersion is the considered factor. The above figure 2.1 presents the scenario of BER related in OWC.

For reducing the probability of loss of information LOW density parity check code (LDPC) is an effective error correcting code that used in noisy communication channel. By employing LDPC this probability can be reduced to as small as desired thus the data transmission can be as close to Shannon's limit.

Optical wireless communication is a dynamic area of research and development. The basic scheme of OWC systems communications and applications of are discussed in [4]. With the main focus on the indoor deployment scenarios, here reviewing and summarizing recent advancements in OW communication [5].

For future high performance and cost-effectiveness of indoor optical wireless systems and possible configurations for modulation, and multi-access techniques are presented in [6].

An overview of both operate term in the short- (personal and indoor systems) and the long-range (outdoor and hybrid) regimes and discussed within the context of point-to-point ultra-high-speed data transfer. The discussion on outdoor systems focuses on the impact of atmospheric effects on the optical channel and associated mitigation techniques that extend the realizable link lengths and transfer rates. [7]. The authors have discussed about the generic propagation models for the design of wireless optical communication in [5]. Bit error rate of optical wireless communication had addressed by employing a detector array in the presence of turbulence [8]. Pervasive work and study have done to improve performance of the OWC systems through BER by occupying several methods. These articles focused and discussed on the BER of wireless optical communication presenting different methods and approaches [9] [10] [11].

The artificial neural network (ANN) has some attractive properties like adaptability, parallel processing, and universal approximation. A well train ANN can perform the task equalization efficiently. In this paper authors have presented the ANN for equalization in indoor environment in an optical channel [12]. Another work has done where authors described their research about the improvement of the optical wireless system using Neuro-Fuzzy based Spot-Diffusing Techniques which provides better performance compared to other generic spot-beam diffusion method [13]. However, in optical wireless communications LDPC method is effective way to improve the BER (bit error rate).

Many works have done focused on the LDPC for optical wireless communication. This paper investigated the performance of regular and irregular, random like and structured generalized low-density parity-check (GLDPC) codes for long-haul transmission [14]. The authors evaluated the low-density parity check (LDPC) coded schemes with unequal transmission power allocation (UTPA) in optical wireless channel [15]. This work proposed a method of establishing a secure and reliable communication link using optical wireless communication (OWC) [16]. Here implemented a modified Low-Density Parity-Check (LDPC) codec algorithm in ultraviolet (UV) communication system [17]. [18] Here authors analysis the BER of Optical Wireless Communication (OWC) System employing Neuro-Fuzzy (NF) based spot diffusion system. But this system has not yet implemented for the performance of BER using LDPC. However, in recent years the researches of LDPC for indoor optical wireless communication are increasing day by day.

Now comparing the proposed work where introduce the LDPC methods focusing on the improvement of bit error rate analysis for more secure and reliable data transmission in indoor optical wireless communication area.

III. SYSTEM MODEL & PROPOSED METHOD

In this chapter, an established Spot Beam algorithm using LDPC has been proposed. Here, SNIR, Delay Spread, BER has been considered. This research is worked on the existing Spot Beam Selection algorithm using LDPC methods for high speed data transfer and reduce the probability of loss of information.

An Indoor Optical Wireless System: RF (Radio Frequency) and Infrared are two major transmission technologies to gain indoor optical wireless communication [19]. Its performance depends on the propagation and type of system used. The basic systems types can be categorized into diffuse or line of sight (LOS) systems [6], [7]. To obtain high data rates such as Gbit/s can be achieved in LOS systems, [8], [11], but due to its directionality the system is vulnerable to blockage/shadowing. Whereas several paths from source to receiver exist in diffuse OW system, which makes the system robust to blockage/shadowing. However, the path losses are high and multi-paths create inter-symbol

interference (ISI) which limits the achievable data rate [8] [11].

Transmission Techniques: Basic optical wireless system consists of a transmitter (using LEDs or LDs). Propagation medium like free space and the receiver (using APDs or PIN diodes). Some transmission techniques are described below:

Directed beam infrared (DBIR) radiation: From the transmitter to the receiver the optical beam travels directly without any reflection. The optical wireless link using this technique is established between two fixed data terminals with highly directional transmitter and receiver at both ends of the link. Lack of mobility and susceptibility to blocking and shadowing by personnel and machines is the main drawback of this technique.

Diffuse infrared (DFIR) radiation: The transmitters send optical signals in a wide angle to the ceiling and after one or several reflections the signals arrive at the receivers in DFIR. For transmission, the system does not require any line of sight alignment which is one of the most desirable configurations from a user point of view. Though, systems using this technique have a higher path loss than their DBIR counterparts, requiring higher transmitter power levels and receivers with larger light collection area. Multipath dispersion is another challenging problem in this technique. Radiate optical power over a wide solid angle. Thus, provide mobility to the receiver.

Quasi-diffuse infrared (QDIR) radiation: In QDIR, there is a base station (BS) with a relatively broad coverage made of passive or active reflector which usually accumulated on the ceiling. By always maintaining the line of sight BS transmits (receives) the signal power to (from) the remote terminals (RTs) thus, the RTs cannot be fully mobile. From any position in the room to enable communication between itself and the BS the RT's transceiver must be aimed to the BS, or its FOV must be wide enough. In another appearance of QDIR technique, the transmitter may send the optical signal to a designated area on the ceiling and the receiver is supposed to face that area. In general, this architecture provides a concession between the DFIR and DBIR option. In QDIR system inherits aspects of both point-to-point and diffuse links. Gradually deviating beam sources which illuminate a grid of spots on the ceiling.

Channel propagation: A number of transmission techniques are possible for indoor optical wireless systems; these techniques may be classified according to the degree of directionality of transmitter and receiver [19]. For indoor OW links, the two most common configurations are LOS and non-LOS propagation systems. In the field of indoor mobile application LOS links rely upon a direct path between the transmitter and receiver, regardless of their beam angles, while non-LOS links generally rely upon light reflections from walls, ceilings and other diffuse reflecting surfaces a non-directed non-LOS link scenario which is often referred as a diffuse link is the most desirable approach. In the environments where shadowing exists diffuse systems also play a very significant role. One of the main reasons is the facts that diffuse propagation systems do not require transmitter-receiver alignment or line-of-sight and instead make use of reflections from walls, ceiling, and other reflectors. However, diffuse systems are subject to multipath dispersion which results in signal spread and inter symbol interference. Through the use of diversity and/or equalization the effects of multipath propagation can be reduced.

System Model: Here we consider an empty room with floor dimensions of $8 \times 4 \text{ m}^2$ ceiling height of 3m. Where the reflection coefficient of the ceiling is considered to be 0.8. The ceiling has eight spot lights. In the Figure x_0 and x are the position of the imaging receiver and v is the velocity where α is elevation angle δ , is the azimuth angle, $d = 8$, $w = 4$ and $h = 3$. Neuro-Fuzzy (NF) adaptive multibeam

transmitter is located at the center of the room whereas an imaging receiver is placed at

$$\mathbf{x}_0 = (1; 1; 0; 5).$$

With adapted beam angle (α, δ) and power that is reflected by the imaging receiver while transmitter generates multi spot beam matrix on the ceiling. Through the low rate diffuse channel, the transmitter learns receiver position, mobility. At low data rate, the beam maintains the fixed power.

Analysis of SNIR and BER: The ambient light affects signal-to-noise-plus interference (SNIR) at the receiver in indoor optical-wireless communication. Many researchers have considered intensity modulation with direct detection (IM/DD) as most feasible approximation. The received signal, symbolize by $\mathbf{y}(t)$ can be expressed as:

$$\mathbf{y}(t) = \sum R\mathbf{x}(t) * \mathbf{h}(t, \alpha, \delta) + \sum \mathbf{n}(t, \alpha, \delta) + \sum I(t, \alpha, \delta) \dots \dots (1)$$

Where R is the receiver responsivity, $\mathbf{x}(t)$ is the instantaneous optical transmitted power, $\mathbf{h}(t, \alpha, \delta)$ is the impulse response of the OW channel, $\mathbf{n}(t, \alpha, \delta)$ is the ambient light noise, $I(t, \alpha, \delta)$ is the instantaneous interference power. The SNIR, denoted by γ , of the received signal can be calculated by [9].

$$\gamma = \frac{R^2(P_{s1}-P_{s0}) \times h^2}{(\sigma_{s1}-\sigma_{s0})^2} \dots \dots (2)$$

Where P_{s1} and P_{s0} are the optical power associated with the binary 1 and binary 0 respectively, σ_{s1} σ_{s0} are the shot noise variation component with P_{s1} and P_{s0} respectively.

Bit Error Rate: For the non-encoded system with binary phase-shift-keying(BPSK), the BER expression can be given by:

$$\psi_{bpsk}(\gamma) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{b_{bpsk}\gamma}{\sin^2\varphi}\right) \dots \dots (3)$$

Where $b_{bpsk} = \sin^2(\pi/2)$. By (2) and (3) we can write

$$\psi_{bpsk}(\gamma) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{\sin^2(\pi/2) \left(\frac{R^2(P_{s1}-P_{s0}) \times h^2}{(\sigma_{s1}-\sigma_{s0})^2}\right)}{\sin^2\varphi}\right) \dots (4)$$

Adaptive Power Allocation: The achievable data transmission rate, denoted by b , of the OWC system is given by $b = \frac{1}{M} \sum_{i=1}^M \log_2 \left(1 + \frac{R^2 \times ((P_{s11}-P_{s01})h_i^2)}{(\sigma_{s11}-\sigma_{s01})^2}\right) \dots (5)$

The optimization problem and constraint of the power allocation can be written as

$$\max \quad b \dots \dots (6)$$

$$\text{s.t} \quad \sum_{j=1}^J P_j \leq \bar{P} \dots \dots (7)$$

Where \bar{P} is the average power. To analyze the above optimization problem, we can use the Lagrange multiplier method and the Lagrangian function is defined as $L = b + \mu_j (\sum_{j=1}^J P_j \leq \bar{P}) \dots \dots (8)$

where μ_j is the Lagrange multiplier. After solving the Eqn. (8), we can write

$$P_j = \left[\frac{(P + \sum_{j=1}^J \frac{1}{h_i})}{c} - \frac{1}{h_i} \right] \dots \dots (9)$$

$$= \max \left[\lambda(C) - \frac{1}{h_i}, 0 \right] \dots \dots (10)$$

Delay Spread: The Doppler spread of an impulse is expressed as rms value by,

$$D = \sqrt{\frac{\sum (t_i - \mu)^2 P_i^2}{P_i^2}} \dots \dots (11)$$

Where $\mu = \frac{t_i P_i^2}{P_i^2}$ and t_i is the delay time and P_i is the received power. Now considering the Doppler shift and the Neuro fuzzy ANFIS model and using the algorithm for Spot Beam selection of this paper.

Proposed System Method: Since the aim of this proposed work is to transmit correct information at longer distance with lower transmit power, LDPC coding scheme is used with coded modulation to achieve significant coding gain without bandwidth expansion.

System model can be explained by using LDPC where, information typically in the form of digital data, is input to electronic circuitry that modulates the transmitting light source (LEDs/LDs). The source output passes through an optical system into the free space (propagation medium). The received signal also comes through the optical system and passes along the optical signal. The optical signal traverses through the indoor channel where it gets affected by the multipath propagation and back ground noise. The noise affected optical signal is detected by the photo detector.

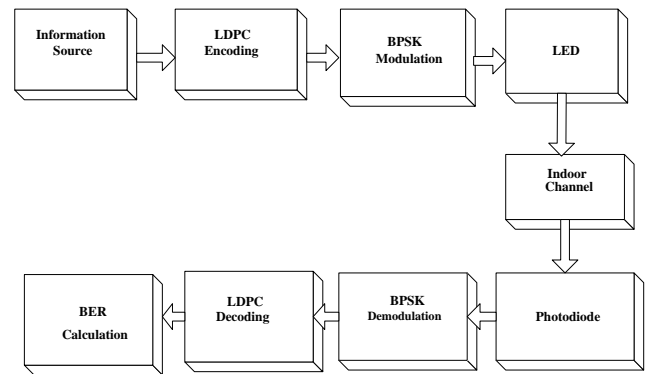


Fig. 3.1: Block Diagram for the LED based Indoor Communication System Model using LDPC coding with BPSK Modulation

Then BPSK demodulation takes place followed by LDPC decoding. The BER analysis is done to estimate the performance of the system.

Error Correction Coding Scheme – LDPC: LDPC codes are based on a parity-check matrix where each transmission message contains (m, n) binary block with 2^n binary m tuples $(m > n)$. Here m is block length, n is number of data bits and $(m - n)$ is number of checked bits. Code rate, $\epsilon = \frac{n}{m}$

Encoding: The Parity-check matrix, H , for (7, 4) LDPC block code is constructed using hamming codes as constituent codes so that the position where error occurs can be detected for correction.

$$H = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \dots \dots (12)$$

With the associated data bits of parity locations, Generator matrix, G , is generated and given below:

$$G = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \dots \dots \dots (13)$$

Given a message u , the codeword c will be the product of G and u .

$$c = G.u \dots \dots \dots (14)$$

If the received codeword is v , the syndrome vector, denoted by z , will be

$$z = H.v \dots \dots \dots (15)$$

When $z = 0$, codeword is error free. Otherwise the value of z is the position of the flipped bit.

Decoding: By sending messages through the connected edges between variable nodes $b_i \in B$ and check nodes $f_j \in F$ of factor graph where B and F are the set of all message nodes and check nodes respectively decoding is accomplished. Here assume that $m_{f_j b_i}$ is the message sent by node b_i to $m_{f_j b_i}$ vice-versa. Let, the probability that there is a 1 at each variable node b_i is α_i where $i = 1, 2, 3, \dots n$. Now, the probability that there is an even number of 1's on variable nodes is,

$$P_{\text{even}} = P_r [b_1 \oplus \dots \oplus b_n] \\ = \frac{1}{2} [1 + \prod_{i=1}^n (1 - 2\alpha_{b_i})] \dots \dots \dots (16)$$

Now, $m_{f_j b_i}(0)$ can be expressed as,

$$m_{f_j b_i}(0) = \frac{1}{2} \left[1 + \prod_{b' \in B \setminus b_i} (1 - 2\alpha_{b'}) \right] \dots \dots (17)$$

$$\text{Therefore, } m_{f_j b_i}(1) = 1 - m_{f_j b_i}(0) \dots \dots \dots (18)$$

The variable nodes update their response messages to check nodes using equation (18).

$$m_{f_j b_i}(0) = k_{b_i f_j} \left[(1 - P_i) \prod_{f' \in F \setminus f_j} (m_{f' b_i}(0)) \right] \dots \dots (19)$$

And

$$m_{f_j b_i}(1) = k_{b_i f_j} \left[P_i \prod_{f' \in F \setminus f_j} (m_{f' b_i}(1)) \right] \dots \dots (20)$$

Here $k_{b_i f_j}$ is the constant to ensure

$$m_{f_j b_i}(0) + m_{f_j b_i}(1) = 1$$

$$P_i = P_r [b_i = 1|v] \dots \dots \dots (21)$$

At this point, b_i nodes update their decision, denoted by ζb_i , by calculating the probabilities $Mb_i(0)$ and $Mb_i(1)$ for 0 and 1 respectively.

$$Mb_i(0) = k b_i (1 - P_i) \prod_{f_j \in F} m_{f_j b_i}(0) \quad \text{and} \quad Mb_i(1) \\ = k b_i (P_i) \prod_{f_j \in F} m_{f_j b_i}(1).$$

Therefore, decision, denoted by ζb_i is given below:

$$\zeta b_i = \{1 \text{ if } Mb_i(1) > Mb_i(0) \text{ else.}\}$$

The algorithm will terminate if ζb_i satisfies parity-check equation, otherwise predetermined number of iteration will terminate it. Now Algorithm for correct data transmission is summarized as follows: A spot beam scans the ceiling, SNIR and delay spread, $\Delta \sigma$ for each beam have been calculated by the image receiver using Eqs (2) and (9). Based on the required minimum SNIR, i.e., min, and maximum delay spread, i.e., max, transmitter selects the spot-beam matrix (H) by NF controller. The transmitter allocates the power for each selected beam adaptively using Eqn (7). Based on Doppler shift, the transmitter adapts the beam angles α and δ . Multi-spot optical transmitter further reduces the by scheduling. Finally, Multi-spot optical transmitter transmits the spot beam matrix to receiver via ceiling. When transmitter achieves the receiver current position then it transmits data using LDPC; after getting the data on the other hand decode the data using decoding LDPC code. Go to Step 1 if transmitter gets receiver's position update.

Algorithm: The following algorithm will find the spot beam with an equal power allocation over 40x20 beam hologram or matrix, H.

Notation:

γ =SNIR, σ = delay spread, α = azimuth angle and δ = elevation angle, T=transmitter, R=Receiver.

P=position of receiver.

Initialization;

Pro1: begin

for each beam i=1 to n do

Calculates γ and σ ;

Set γ_{\min} and σ_{\max} ;

if $\gamma \leq \gamma_{\min}$ and $\sigma \geq \sigma_{\max}$ then

Calculates α and δ ;

end

Calculate P using ANFIS Controller;

T sends encoded data c using LDPC to R located at position P ;

R decoded data c' and receives it;

end

end

if P of R changes then

Call Pro1

End

IV. NUMERICAL ANALYSIS

The proposed method is worked on the implanted Neuro-Fuzzy based multi beam system (NFMS) diversity receiver configuration. It has been also compared with other spot-beam diffusion method. The research has been done by employing LDPC code into the above investigated method. Thus, Performance for the indoor optical wireless communication improves and shows possible low error rate in the purpose of data transmission.

Result and Discussion: For the analysis of the implemented algorithm, the simulation parameters those are considered and implanted: length, width and height of the room are 8m, 4m and 3 m; the reflection coefficient of the ceiling is $\rho = 0.8$. One transmitter which is located at (2; 4; 1) location; there is also one receiver; the area acceptance semi-angle of each photo-diode are 2cm^2 and 65° respectively. The number of pixel at the receiver is 200 (with area of 0.01cm^2) Pedestrians move typically at the speed of 1 m/s. If the SNIR is computed after $10 \mu\text{s}$; there are 8 spot lamp in the room which are located at (1; 1; 1), (1; 3; 1), (1; 5; 1), (1; 7; 1), (3; 1; 1), (3; 3; 1), (3; 5; 1), and (3; 7; 1); and

the wavelength of the light is 850nm. The results of the Adaptive Spot Beam Selection applying ANFIS model have been implanted by using this parameter.

In this section, the numerical results are analyzed using the mathematical equations derived in the previous Chapter and are simulated using the codes written in MATLAB. The performance analysis is based on a scenario for a typical indoor environment in the presence of additive white Gaussian noise. The performance of the system depends on the receiver sensitivity. The 80 ms adaptation time will give overhead of 8%. Adaptation time depends on environment. Receiver computes the SNIR and delay spread and sends this information via a low rate channel to the transmitter. ANFIS consider two inputs. Iterative training of the ANFIS has been done to achieve the desired output. After a predefined simulation time to obtain the simulation result and use them to train.

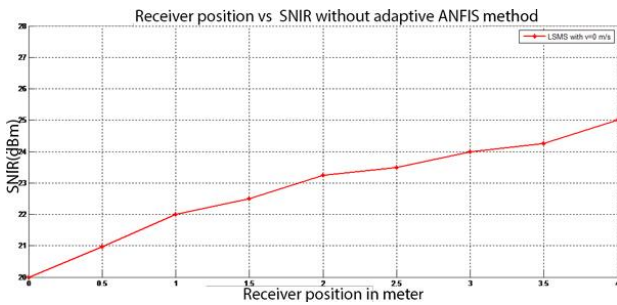


Figure 4.1: Effect of receiver position on SNIR distribution for without Adaptive Neuro Fuzzy System

Another approach is placed here for the Adaptive method is shown in the figure below:

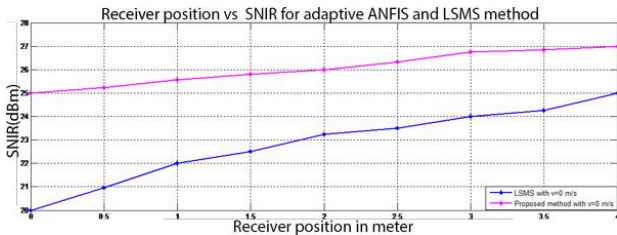


Figure 4.2: Effect of receiver position on SNIR distribution for Adaptive ANFIS Method and LSMS

From this figure, it seems that here indicates a comparison between proposed model of Adaptive Neuro Fuzzy Method and line strip multi-spot diffuse system (LSMS). The BER calculations were performed for the receiver moving towards the transmitter (i.e., the value of x is increasing) while neglecting the movement along y axis. Significant BER improvement is observed as the NFC moves the spot beam, selects the best positioned spot only, and allocate the power adaptively based on the channel condition of the selected slots. It is also found that the BER performances have been degraded as the receiver is moving. The BER increases as the velocity of the receiver increases.

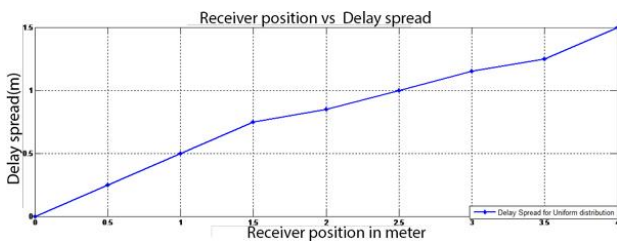


Fig 4.3: Effect of receiver position in Delay Spread

This figure shows the BER comparison for proposed model with LSMS and conventional diffuse system. Performance evaluation reveals that BER has been improved, if we change the spot beam angle adaptively.

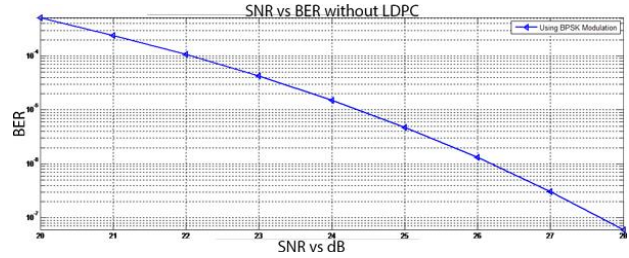


Fig 4.4: SNIR and BER Using BPSK Modulation

This figure described the SNIR and BER performance Using BPSK Modulation technique.

This shows the BER changes in response to the SNIR.



Figure 4.5 represents effects of SNIR and BER using LDPC. Where it indicates the BER reduce in response to the SNIR.

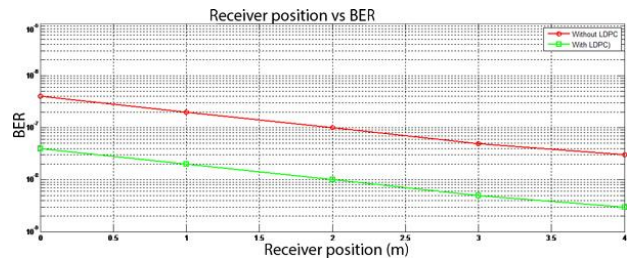


Fig 4.6: BER for Corresponding Receiver position

From the figure 4.6 it has been established that while using LDPC coding scheme then it reflects the better performance. This graph describes the comparison between the performance for the non-coded system and coded (LDPC) system. In which BER consistently increases. Where comparatively LDPC Gives enough pretty result that the proposed method expected.

So, from the above simulation and analysis the study for the desired result has been investigated. Thus, it describes that the Adaptive Spot Diffusion Beam Selection algorithm if coded with LDPC then generates better BER performance.

VII. CONCLUSION AND RECOMMENDATIONS

It is possible to conclude that, in spite of the advances achieved so far, there is still a lot of work to be done to exploit completely the advantages and the potential offered by the optical medium. For indoor wireless system applications, the use of optical communication offers an important alternative or the growing area of mobile computers and communication. Thus, techniques to improve the operation and speeds of infrared wireless systems with room environments have still to found, while trying to decrease the cost of the systems as much as possible. Researcher and manufacturers are also trying to find ways to improve the data bit rates and the range offered by current systems. This work evaluates the performance of a LDPC coded for indoor optical wireless communication system to increase the low bit error rate which can be used for secure and reliable response communication where quality of service is the main concern. Simulation results indicated that BER performance has been improved by incorporating LDPC coding. This work can be extended by considering the mobility and BER of multiple transmitters and receivers using this coding scheme.

REFERENCES

- [1] Nagatsuma T., T. Tkakda, H. J. Song, K. Ajito, N. Kukutsu, and Y. Kado, "Millimeter- and THz-wave photonics towards 100-Gbit/s wireless transmission," *Proc. of IEEE Photonics Society's 23th Annual Meeting*, 385-386, Denver, Colorado, USA, 2010.
- [2] F.R. Gfeller and U. H. Bapst, "Wireless in-house data communication via diffuse infrared radiation", *Proc. IEEE*, vol. 67, no. 11, pp.1474-1486. Nov. 1979.
- [3] T. Komine, M. Nakagawa, "A Study of Shadowing on Indoor Visible-Light Wireless Communication Utilizing Plural White LED Lightings. *Int. Sympo. On Wireless Commun*", pp. 36-40, 2004.
- [4] J. Sánchez-López, A. A. M, F. J. Mendieta and Iván Nieto Hipólito (2011). "Trends of the Optical Wireless Communications, *Advanced Trends in Wireless Communications*", Dr. Mutamed Khatib (Ed.), ISBN: 978-953-307-183-1.
- [5] Z. Ghassemlooy and A. R. Hayes, "Indoor Optical Wireless Networks", *IEEE Commun. Mag.*, vol. 36, no. 12, pp. 88-94, December 2011.
- [6] C. Sing, J. John, Y.N.Singh and K.K.Tripathi, "A Review on Indoor Optical Wireless Systems", *IEEE Commun. Mag.*, vol. 33, no. 15, pp. 79-84, December 2012.
- [7] Z. Zhang, L. Dolecek, B. Nikolić, V. Anantharam and M. J. Wainwright "Design of LDPC Decoders for Low Error Rate Performance", *IEEE Commun. Mag.*, vol. 32, no. 13, pp.57-69, June 2014.
- [8] O. Bouchet, M. Bertrand, P. Besnard, "Optical Wireless Communication: LOS/WLOS/DIF propagation model and QOFI software", *IEEE Commun. Mag.*, vol. 35, no. 17, pp.69-74, December 2010.
- [9] Dang A, "A closed-form solution of the bit-error rate for optical wireless communication systems over atmospheric turbulence channels", 2011 Feb 14; 19(4):3494-502.doi: 10.1364/OE.19.003494.
- [10] Aharonovich M, Arnon S, "Performance improvement of optical wireless communication through fog with a decision feedback equalizer", *Opt Soc Am A Opt Image Sci Vis*. 2005 Aug; 22 (8):1646-54.
- [11] Z. Ghassemlooy, C. K. See and J. M. Holding and C. Lu, "Bit-error-rate analysis for hybrid PIM-CDMA optical wireless communication systems", 16 AUG 2001 DOI: 10.1002/mop.1351.
- [12] A. K. Jaiswal, Anil Kumar, Santosh Tripathi, A. K. C. Shiats, "To Study the Effect of BER and Q-factor in Intersatellite Optical Wireless Communication System", *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)* ISSN: 2278-2834, ISBN: 2278-8735. Volume 3, Issue 4 (Sep-Oct. 2012), PP 19-21.
- [13] S. Rajbhandari, Z. Ghassemlooy and M. Angelova, "Performance of OOK with ANN Equalization in Indoor Optical Wireless Communication System", *IEEE Commun. Mag.*, vol. 32, no. 14, pp.55-65, March 2010.
- [14] I. B. Djordjevic, O. M. and B. Vasic, "Generalized Low-Density Parity-Check Codes for Optical Communication Systems", *IEEE Commun. Mag.*, vol. 37, no. 15, pp.85-95, April 2013.
- [15] Sumikawa K. and Habuchi H., "Optical wireless LDPC-coded PPM with unequal transmission power allocation scheme Oct. 27 2013-Nov. 1 2013 Page(s):44 - 47Print ISBN:978-1-4799-6028-6 INSPEC Accession Number:14431132 Conference Location :Tokyo DOI:10.1109/IWSDA.2013.6849058 Publisher:IEEE.
- [16] M. Saadi, A. Bajpai, Y. Zhao, P. Sangwongngam and L. Wuttisittikulkij, "Design and Implementation of Secure and Reliable Communication using Optical Wireless Communication", *Frequenz* 2014; 68(11-12): 501-509 DOI 10.1515/freq-2014-0027 Received February 24, 2014.
- [17] M. Wu, D. Han, X. Zhang, F. Zhang, M. Zhang, and G. Yue, "Experimental research and comparison of LDPC and RS channel coding in ultraviolet communication systems", published 28 Feb 2014 (C) 2014 OSA 10 March 2014, Vol. 22, No. 5, OI:10.1364/OE.22.005422, *OPTICS EXPRESS* 5423.
- [18] S. A. Mamun, M. S. Kaiser, M. R. Ahmedz and Md. I. Islam, "BER Analysis of Optical Wireless Communication System Employing Neuro-Fuzzy based Spot-Diffusing Techniques". "*International Conference on Electrical Information and Communication Technology (EICT)*, 2013", ISBN: 978-1-4799-2297-0, DOI: 10.1109/EICT.2014.6777842, Khulna, IEEE.
- [19] J. Fernandes, P. A. Watson and J. Neves, "Wireless LANs : Physical properties of Infrared Systems vs MmW Systems", *IEEE Commun. Magazine*, August 1994.

Risk Management Process Analysis for Information & Communication Technology (ICT) Systems: Risk Mitigation perspective

Nwagu, Chikezie Kenneth, Omankwu, Obinnaya Chinecherem, Okonkwo, Obikwelu Raphael

¹**Computer Science Department, Nnamdi Azikiwe University Awka Anambra State, Nigeria.
Nwaguchikeziekeneth@hotmail.com**

²**Computer Science Department, Michael Okpara University of Agriculture Umudike Umuahia, Abia State,
Nigeria.
Saintbeloved@yahoo.com**

³**Computer Science Department, Nnamdi Azikiwe University Awka Anambra State, Nigeria.
Oobi2971@yahoo.com**

Abstract— *This Research Paper, explored Risk management from Risk mitigation perspective. Risk Mitigation Perspective thoroughly examined mitigation processes, Strategies for mitigating both Negative and Positive Risks, discussed essence of Cost-benefit analysis for each recommended control as core mitigation process in order to select cost-effective and best-fit control for the Organization and stakeholders. It further emphasises importance of integration and updating of risk registers in the process and consistent communication/consultation among stakeholders including the Original Equipment Manufacturers (OEMs).*

Keywords—*Risk, Risk Management, Risk Mitigation, cost-benefit analysis, Vulnerability, Opportunity*

I. INTRODUCTION

In the face of global ever-dynamic threats and attacks, every Organization is adopting measures to reduce negative risks and utilize positive risks. This ensures that her vision and mission is protected, guarded and fully enhanced. This critical as Organizations make ICT, hub for better support and sustenance of her business. As Organizations automated their processes using data and communication devices, Risk Management plays a very critical role in protecting the organizations information assets, and therefore its mission. Risk Management is every stakeholder`s duty and not only for the technical IT team. Therefore, it should be treated as fundamentally as an essential role of the Management. An effective and efficient risk management process is an important component of a successful ICT security so as to ensure data confidentiality, integrity and high availability. According to Project Management Institute (PMI) (2015), Risk is an uncertain event or condition that, if it occurs, has a positive or negative effect on one or more project objectives such as scope, schedule, cost, and quality. A risk may have one or more causes and if it occurs, it may have one or more impacts. **Risk management** is the process of identifying risk, assessing risk, and taking steps to reduce risk to an acceptable level, if possible eradicate it completely. Risk should be both net negative effect of exercise of vulnerability (weakness) according to stoneburner et al (2002) and net positive effect of harness of addendum opportunity (Strength) in the devices in order to fully maximize their utilization and functionality

II. RISK MITIGATION PROCESS

This step involves process of weighing and developing options and actions to enhance opportunities and reduce threats to the devices. Risk mitigation, as the second process of risk management, involves prioritizing, evaluating, and implementing the appropriate risk-reducing or risk-enhancing controls recommended from the risk assessment process. Because the elimination of all risk is usually impractical or close to impossible, it is the responsibility of principal stakeholders to use the least-cost approach and implement the most appropriate controls to decrease mission risk to an acceptable level, with minimal adverse impact on the organization`s resources and mission. Risk mitigation options, an

approach for control implementation, control categories, the cost-benefit analysis used to justify the implementation of the recommended controls, and residual risk. Of course, updating of appropriate risk registers.

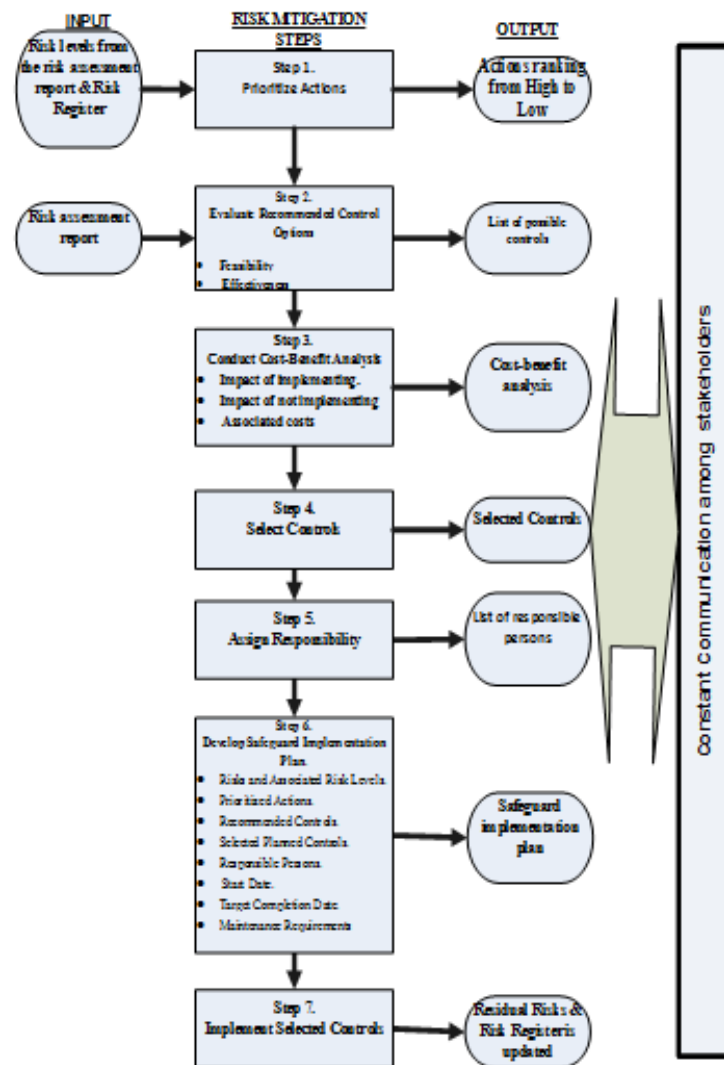


Figure 1: Risk Mitigation Process (adapted from stoneburner et al)

Step 1: Prioritize Actions

Based on the risk levels presented in the risk assessment report and content of risk register, the recommended implementation actions are prioritized. In allocating resources, top priority should be given risk items with unacceptably high risk rankings (e.g., risk assigned a Very High or High risk level). These vulnerability/threat pairs will require immediate corrective action to protect an organization's interest and mission.

Step 2: Evaluate Recommended Control Options

The controls recommended in the risk assessment process may not be the most appropriate and feasible options for a specific organization and IT system. During this step, the feasibility (e.g., compatibility, user acceptance) and effectiveness (e.g., degree of protection and level of risk mitigation) of the recommended control options are analyzed. The objective is to select the most appropriate control option for minimizing risk.

Step 3: Conduct Cost-Benefit Analysis

According to Wethli, K. (2014), Cost-Benefit analysis provides one means of identifying the cases in which specific interventions to manage risk do appear to be cost-effective.

To aid management and principal stakeholders in decision-making and to identify cost-effective controls, a cost-benefit analysis is conducted. The objectives and method of conducting the cost-benefit analysis will be discussed in detail later in this work.

Step 4: Select Control

On the basis of the results of the cost-benefit analysis, management determines the most cost-effective control(s) for reducing risk to the organization's mission. The controls selected should combine technical, operational, and management control elements to ensure adequate security for the IT system and the organization.

Step 5: Assign Responsibility

Appropriate stakeholder (in-house personnel or external contracting staff) who have the appropriate expertise and skill-sets to implement the selected control are identified, and responsibility is assigned by email alert. However, basic user awareness campaign must be carried out company-wide to educate users on what to do and how to do it, including sharing the in-house escalation matrix. All these are to ensure total proactivity in mitigating risks.

Step 6: Develop a Safeguard Implementation Plan

During this step, a safeguard implementation plan (or action plan) is developed. The plan should, at a minimum, contain the following information:

- a) Risks (vulnerability/threat pairs) and associated risk levels (output from risk assessment report)
- b) Recommended controls (output from risk assessment report)
- c) Prioritized actions (with priority given to items with Very High and High risk levels)
- d) Selected planned controls (determined on the basis of feasibility, effectiveness, benefits to the organization, and cost)
- e) Required resources for implementing the selected planned controls
- f) Lists of responsible teams and staff
- g) Start date for implementation
- h) Target completion date for implementation
- i) Maintenance requirements.

The safeguard implementation plan prioritizes the implementation actions and projects the start and target completion dates. This plan will aid and expedite the risk mitigation process.

Step 7: Implement Selected Control(s)

Depending on individual situations, the implemented controls may lower the risk level and may not eliminate the risk completely. This gives room for Residual risks which are kept under close watch and monitoring in the appropriate risk registers

Control Categories

In implementing recommended controls to mitigate risk, an organization should consider technical, management, and operational security controls, or a combination of such controls, to maximize the effectiveness of controls for their IT systems and organization. Security controls, when used appropriately, can prevent, limit, or deter threat-source damage to an organization's mission. The control recommendation process will involve choosing among a combination of technical, management, and operational controls for improving the organization's security posture. The trade-offs that an organization will have to consider are illustrated by viewing the decisions involved in enforcing use of complex user passwords to minimize password guessing and cracking. In this case, a technical control requiring add-on security software may be more complex and expensive than a procedural control, but the technical control is likely to be more effective because the enforcement is automated by the system.

Residual Risks and Updating of Risk Register

As the final output of risk mitigation process, Organizations can analyze the extent of the risk reduction generated by the new or enhanced (recommended) controls in terms of the reduced threat likelihood of occurrence or impact - the two parameters that define the mitigated level of risk to the organizational mission. And Of course, update the risk register either by including list of risks that have least impact when all are compared and/or de-listing the ones which have been triggered and considered of high impact and invariably mitigated.

Implementation of new or enhanced controls can mitigate risk by:

- a) Eliminating some of the system's vulnerabilities (flaws and weakness), thereby reducing the number of possible threat-source/vulnerability pairs or outright prevention of exercise of the vulnerabilities
- b) Adding a targeted control to reduce the capacity and motivation of a threat-source .For example, a department determines that the cost for installing and maintaining add-on security software for the stand-alone PC that stores its sensitive files is not justifiable, but that administrative and physical controls should be implemented to make physical access to that PC more difficult (e.g., store the PC in a locked room, with the key kept by the manager).
- c) Reducing the magnitude of the adverse impact (for example, limiting the extent of vulnerability or modifying the nature of the relationship between the IT system and the organization's mission).

III RISK MITIGATION OPTIONS FOR NEGATIVE AND POSITIVE RISKS RESPECTIVELY

These are various options through which threats can be reduced and where possible, eradicated (in case of negative risks). However, in case of positive risks - opportunities, they can be enhanced or invested upon. Based on these two sides of a risk, its mitigation options are treated in these lights as well

Strategies for Negative Risks or threats: There are three main strategies used to deal with threats that may lead to compromise of data/information integrity and confidentiality by exploiting the vulnerability in the devices; if they occur are:

- a) Risk Avoidance: This is used where the risk impact is high. The stakeholders act to eliminate the threats .The most radical avoidance strategy is to shut down the devices or disconnect them from network. This may prompt the stakeholders to consult the manufacturers for immediate solution, if there is no other alternative.
- b) Risk Transfer: Here, the stakeholders shift the impact of the threat to a third party and ownership of the responsibility by use of insurance, warranties, guarantees etc.
- c) Risk Mitigate: In this strategy, stakeholders act early to reduce the probability of occurrence or impact of a risk. Thereby making the risk to be within acceptable threshold.
- d) Risk Acceptance: This is used for Negative and Positive risks. In this scenario, stakeholders decide to acknowledge the risks and take no action unless the risk occurs. However this strategy provides room for periodic reviews of the threats to ensure that the risk does not change significantly. This also happens for the risks under close monitoring such as those in the risk registers.

Strategies for Positive Risks or Opportunities:

- a) Exploit: This is used for risks with positive impacts on the devices where the stakeholders wish to ensure the opportunity is realized. It seeks to eliminate the uncertainty associated with a particular upside risk by ensuring the opportunity definitely happens. For example, engaging a vast expert to administer the devices who ensures that all the devices` full potential are utilized and also embraces trends of new technologies including their upgrades in order to proactively minimized any vulnerability and negative risks.

- b) Enhance: This is used to increase probability and/or positive impacts of an opportunity. Identifying and Maximizing key drivers of this positive-impact risk may increase the probability of their occurrence. For example, Changing/upgrading the software (Operating systems, application etc.) and hardware of a device will definitely increase the throughput and security.
- c) Share: Sharing a positive risk involves allocating some or all of the ownership of the opportunity to a third party who is best able to capture the opportunity for the benefit of the stakeholders. For example, forming a risk-sharing partnership, teams or joint ventures can be established with express purpose of taking advantage of the opportunity so that all stakeholders gain from their actions.
- d) Accept: Accepting an opportunity is being willing to take advantage of the opportunity if it arises but not practically pursuing it.

The goals and mission of an organization should be considered in selecting any of any of the options. It may not be practical to address all identified risks (low, medium and high), so priority should be given to the risks which adjudged to have potential to cause significant mission impact or harm to the stakeholders and their organizations. Therefore, the best of breed approach is to use appropriate technologies from various vendor security products, along with the appropriate risk mitigation option and other administrative measures which are best practices.

IV. COST-BENEFIT ANALYSIS

To allocate resources and implement cost-effective controls, organizations, after identifying all possible controls and evaluating their feasibility and effectiveness, should conduct a cost-benefit analysis for each recommended control to determine which controls are required and appropriate for their circumstances. High cost of a control does not means most effective and appropriate for a particular risk.

The cost-benefit analysis can be qualitative or quantitative. Its purpose is to demonstrate that the costs of implementing the controls can be justified by the reduction in the level of risk. For example, the organization may not want to spend 1,000NGN on a control to reduce a 200NGN risk.

A cost-benefit analysis for recommended controls encompasses the following:

- a) Determining the impact of implementing the new or enhanced recommended controls
- b) Determining the impact of not implementing the new or enhanced recommended controls
- c) Compatibility and adaptability of the recommended control to the existing ones
- d) Estimating the costs of the implementation. These may include, but are not limited to, the following:
 - i. Hardware and software purchases or upgrades
 - ii. Reduced operational effectiveness if system performance or functionality is reduced for increased security
 - iii. Cost of implementing additional policies and procedures
 - iv. Cost of hiring additional personnel to implement New policies, procedures, or services, if need be.
 - v. Training costs
 - vi. Maintenance costs and /or operational cost if applicable

Assessing the implementation costs and benefits against system and data criticality to determine the importance to the organization of implementing the recommended controls, given their costs and relative impact.

The organization will need to assess the benefits of the controls in terms of maintaining an acceptable mission posture for the organization. Just as there is a cost for implementing a needed control, there is a cost also for not implementing it. By

relating the result of not implementing the control to the mission, organizations can determine whether it is feasible to forgo its implementation. At this stage after considering the above caveats, it is recommended that best-and-purpose fit recommended control be implemented.

V. CONCLUSIONS

Identifying risks is important but mitigating them is much more important, as ensuring safety of stakeholders' investments and of course, mission protection is one of every organization's principal targets. Observing and following the steps as earmarked in the process is key to ensure fulfilment of risk mitigation process. In addition, cost-benefit analysis cannot be overlooked as it is a critical component of the entire process. Hence every recommended control or option must be deeply analyzed to ensure fit for purpose and further assurance of targeted results or impact. In all these, there should be consistent updating of stakeholders and sustenance of the communication throughout life span of the organization. This would further enshrine the process among stakeholders and build organization-wide proactive focus on mitigating high and/or critical medium risks.

VI. REFERENCES

Stoneburner G., Goguen A. and Feringa A. (2001, July). Risk Management Guide for information Systems. Retrieved September 4, 2014 from <http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf>.

PMI. (2012). A Guide to the Project Management Body of Knowledge (PMBOK) Fifth Edition.

Cooney M. (2012, September 21). 10 common mobile Security problems to attack. Retrieved August 22, 2014 from <http://www.pcworld.com/article/2010278/10-common-mobile-security-problems-to-attack.html>

A Guide to the Project Management Body of Knowledge (PMBOK) Fifth Edition.

Siciliano, R. S. (2011). Seven Types of Hacker Motivations. Infosec Island

Risk Management Process Overview. Accessed September 3, 2014 from <http://technet.microsoft.com/en-us/library/cc535304.aspx>

Schneier, B. (2008). Security through Obscurity.

Siciliano, R. S. (2011). Seven Types of Hacker Motivations. Infosec Island

Valsamakis, A. C., (2003). Risk Management.
Heinemann Higher and Further Education (Pty) Ltd. Sandton.

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Dr Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Dr. P. Vasant, University Technology Petronas, Malaysia
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Dr. Praveen Ranjan Srivastava, BITS PILANI, India
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Dr. Tirthankar Gayen, IIT Kharagpur, India
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan
Prof. Ning Xu, Wuhan University of Technology, China
Dr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen
& Universiti Teknologi Malaysia, Malaysia.
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan

Prof. Syed S. Rizvi, University of Bridgeport, USA
Dr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India
Dr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal
Dr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P
Dr. Poonam Garg, Institute of Management Technology, India
Dr. S. Mehta, Inha University, Korea
Dr. Dilip Kumar S.M, Bangalore University, Bangalore
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia
Dr. Saqib Saeed, University of Siegen, Germany
Dr. Pavan Kumar Gorakavi, IPMA-USA [YC]
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India
Dr. J. Komala Lakshmi, SNR Sons College, Computer Science, India
Dr. Muhammad Sohail, KUST, Pakistan
Dr. Manjaiah D.H, Mangalore University, India
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada
Dr. Deepak Laxmi Narasimha, University of Malaya, Malaysia
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India
Dr. M. Azath, Anna University, India
Dr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh
Dr. AOs Alaa Zaidan Ansaef, Multimedia University, Malaysia
Dr. Suresh Jain, Devi Ahilya University, Indore (MP) India,
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia
Dr. Hanumanthappa. J. University of Mysore, India
Dr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)
Dr. Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria
Dr. Santosh K. Pandey, The Institute of Chartered Accountants of India
Dr. P. Vasant, Power Control Optimization, Malaysia
Dr. Petr Ivankov, Automatika - S, Russian Federation
Dr. Utkarsh Seetha, Data Infosys Limited, India
Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal
Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore
Assist. Prof. A. Neela madheswari, Anna university, India
Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India
Mr. Kamanashis Biswas, Daffodil International University, Bangladesh
Dr. Atul Gonsai, Saurashtra University, Gujarat, India
Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand
Mrs. G. Nalini Priya, Anna University, Chennai
Dr. P. Subashini, Avinashilingam University for Women, India
Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat
Mr Jitendra Agrawal, : Rajiv Gandhi Proudlyogiki Vishwavidyalaya, Bhopal
Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India
Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai

Assist. Prof. Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India
Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah
Mr. Nitin Bhatia, DAV College, India
Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India
Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia
Assist. Prof. Sonal Chawla, Panjab University, India
Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India
Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia
Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia
Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India
Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France
Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India
Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of Technology,
Durban, South Africa
Prof. Mydhili K Nair, Visweswaraiah Technological University, Bangalore, India
M. Prabu, Adhiyamaan College of Engineering/Anna University, India
Mr. Swakkhar Shatabda, United International University, Bangladesh
Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan
Mr. H. Abdul Shabeer, I-Nautix Technologies, Chennai, India
Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India
Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran
Mr. Zeashan Hameed Khan, Université de Grenoble, France
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India
Dr. Maslin Masrom, University Technology Malaysia, Malaysia
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City
Dr. Mary Lourde R., BITS-PILANI Dubai , UAE
Dr. Abdul Aziz, University of Central Punjab, Pakistan
Mr. Karan Singh, Gautam Budtha University, India
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia
Assistant Prof. Yasser M. Alginahi, Taibah University, Madinah Munawwarah, KSA
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India
Dr. M.Munir Ahamed Rabbani, B. S. Abdur Rahman University, India
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India
Asst Prof. Jasmine. K. S, R.V.College of Engineering, India
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius
Prof. Mahesh Goyani, G H Patel Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India
Dr. Mana Mohammed, University of Tlemcen, Algeria
Prof. Jatinder Singh, Universal Institution of Engg. & Tech. CHD, India

Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim
Dr. Bin Guo, Institute Telecom SudParis, France
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore
Mr. Milindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India
Dr. C. Arun, Anna University, India
Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology
Subhabrata Barman, Haldia Institute of Technology, West Bengal
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand
Dr. P. Chakrabarti, Sir Padampat Singhania University, Udaipur, India
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India
Mr. Serguei A. Mokhov, Concordia University, Canada
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA
Dr. S. Karthik, SNS College of Technology, India
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain
Mr. A.D.Potgantwar, Pune University, India
Dr. Himanshu Aggarwal, Punjabi University, India
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India
Dr. K.L. Shunmuganathan, R.M.K Engg College , Kavaraipettai ,Chennai
Dr. Prasant Kumar Pattnaik, KIST, India.
Dr. Ch. Aswani Kumar, VIT University, India
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA
Mr. Arun Kumar, Sir Padam Pat Singhania University, Udaipur, Rajasthan
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan

Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA
Mr. R. Jagadeesh Kannan, RMK Engineering College, India
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh
Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia
Mr. R. Mohammad Shafi, Madanapalle Institute of Technology & Science, India
Dr. F. Sagayaraj Francis, Pondicherry Engineering College, India
Dr. Ajay Goel, HIET, Kaithal, India
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India
Mr. Suhas J Manangi, Microsoft India
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded, India
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India
Dr. Amjad Rehman, University Technology Malaysia, Malaysia
Mr. Rachit Garg, L K College, Jalandhar, Punjab
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan
Dr. Thorat S.B., Institute of Technology and Management, India
Mr. Ajay Prasad, Sir Padampat Singhania University, Udaipur, India
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India
Mr. Syed Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh
Mrs Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA
Mr. Anand Kumar, AMC Engineering College, Bangalore
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow, UP India
Dr. V V S S S Balam, Sreenidhi Institute of Science and Technology, India
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India
Prof. Niranjana Reddy, P, KITS, Warangal, India
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India
Dr. A. Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India
Dr. Tossapon Boongoen, Aberystwyth University, UK
Dr. Bilal Alatas, Firat University, Turkey
Assist. Prof. Jyoti Praaksh Singh, Academy of Technology, India
Dr. Ritu Soni, GNG College, India
Dr. Mahendra Kumar, Sagar Institute of Research & Technology, Bhopal, India.
Dr. Binod Kumar, Lakshmi Narayan College of Tech. (LNCT) Bhopal India
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan
Dr. T.C. Manjunath, ATRIA Institute of Tech, India
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan

Assist. Prof. Harmunish Taneja, M. M. University, India
Dr. Chitra Dhawale , SICSR, Model Colony, Pune, India
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India
Mr. G. Appasami, Dr. Pauls Engineering College, India
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan
Mr. Yaser Miaji, University Utara Malaysia, Malaysia
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India
Dr. S. Sasikumar, Roever Engineering College
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India
Mr. Nwaocha Vivian O, National Open University of Nigeria
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India
Mr. Kunal Chadha , R&D Software Engineer, Gemalto, Singapore
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia
Dr. Dhuha Basheer abdullah, Mosul university, Iraq
Mr. S. Audithan, Annamalai University, India
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam
Assist. Prof. Anand Sharma, MITS, Lakshmarangarh, Sikar, Rajasthan, India
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad
Mr. Deepak Gour, Sir Padampat Singhania University, India
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India
Mr. Ali Balador, Islamic Azad University, Iran
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India
Dr. Debojyoti Mitra, Sir padampat Singhania University, India
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia
Mr. Zhao Zhang, City University of Hong Kong, China
Prof. S.P. Setty, A.U. College of Engineering, India
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India
Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India
Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India
Dr. Hanan Elazhary, Electronics Research Institute, Egypt
Dr. Hosam I. Faiq, USM, Malaysia
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India

Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India
Prof Anupam Choudhary, Bhilai School Of Engg.,Bhilai (C.G.),India
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya
Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.
Dr. Kasarapu Ramani, JNT University, Anantapur, India
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India
Dr. C G Ravichandran, R V S College of Engineering and Technology, India
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India
Dr. Nikolai Stoianov, Defense Institute, Bulgaria
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh
Mr. Hemanta Kumar Kalita , TATA Consultancy Services (TCS), India
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys research Group, Venezuela
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia
Dr. Nighat Mir, Effat University, Saudi Arabia
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore
Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India
Mr. P. Sivakumar, Anna university, Chennai, India
Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia
Mr. Nikhil Patrick Lobo, CADES, India
Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India
Assist. Prof. Vishal Bharti, DCE, Gurgaon
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India
Mr. Hamed Taherdoost, Tehran, Iran
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran
Mr. Shantanu Pal, University of Calcutta, India
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria
Mr. P. Mahalingam, Caledonian College of Engineering, Oman
Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt

Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India
Mr. Muhammad Asad, Technical University of Munich, Germany
Mr. AliReza Shams Shafigh, Azad Islamic university, Iran
Prof. S. V. Nagaraj, RMK Engineering College, India
Mr. Ashikali M Hasan, Senior Researcher, CelNet security, India
Dr. Adnan Shahid Khan, University Technology Malaysia, Malaysia
Mr. Prakash Gajanan Burade, Nagpur University/ITM college of engg, Nagpur, India
Dr. Jagdish B.Helonde, Nagpur University/ITM college of engg, Nagpur, India
Professor, Doctor BOUHORMA Mohammed, Univertsity Abdelmalek Essaadi, Morocco
Mr. K. Thirumalaivasan, Pondicherry Engg. College, India
Mr. Umbarkar Anantkumar Janardan, Walchand College of Engineering, India
Mr. Ashish Chaurasia, Gyan Ganga Institute of Technology & Sciences, India
Mr. Sunil Taneja, Kurukshetra University, India
Mr. Fauzi Adi Rafrastara, Dian Nuswantoro University, Indonesia
Dr. Yaduvir Singh, Thapar University, India
Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece
Dr. Vasantha Kalyani David, Avinashilingam University for women, Coimbatore
Dr. Ahmed Mansour Manasrah, Universiti Sains Malaysia, Malaysia
Miss. Nazanin Sadat Kazazi, University Technology Malaysia, Malaysia
Mr. Saeed Rasouli Heikalabad, Islamic Azad University - Tabriz Branch, Iran
Assoc. Prof. Dharendra Mishra, SVKM's NMIMS University, India
Prof. Shapoor Zarei, UAE Inventors Association, UAE
Prof. B.Raja Sarath Kumar, Lenora College of Engineering, India
Dr. Bashir Alam, Jamia millia Islamia, Delhi, India
Prof. Anant J Umbarkar, Walchand College of Engg., India
Assist. Prof. B. Bharathi, Sathyabama University, India
Dr. Fokrul Alom Mazarbhuiya, King Khalid University, Saudi Arabia
Prof. T.S.Jeyali Laseeth, Anna University of Technology, Tirunelveli, India
Dr. M. Balraju, Jawahar Lal Nehru Technological University Hyderabad, India
Dr. Vijayalakshmi M. N., R.V.College of Engineering, Bangalore
Prof. Walid Moudani, Lebanese University, Lebanon
Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India
Associate Prof. Suneet Chaudhary, Dehradun Institute of Technology, India
Associate Prof. Dr. Manuj Darbari, BBD University, India
Ms. Prema Selvaraj, K.S.R College of Arts and Science, India
Assist. Prof. Ms.S.Sasikala, KSR College of Arts & Science, India
Mr. Sukhvinder Singh Deora, NC Institute of Computer Sciences, India
Dr. Abhay Bansal, Amity School of Engineering & Technology, India
Ms. Sumita Mishra, Amity School of Engineering and Technology, India
Professor S. Viswanadha Raju, JNT University Hyderabad, India
Mr. Asghar Shahrzad Khashandarag, Islamic Azad University Tabriz Branch, India
Mr. Manoj Sharma, Panipat Institute of Engg. & Technology, India
Mr. Shakeel Ahmed, King Faisal University, Saudi Arabia
Dr. Mohamed Ali Mahjoub, Institute of Engineer of Monastir, Tunisia
Mr. Adri Jovin J.J., SriGuru Institute of Technology, India
Dr. Sukumar Senthilkumar, Universiti Sains Malaysia, Malaysia

Mr. Rakesh Bharati, Dehradun Institute of Technology Dehradun, India
Mr. Shervan Fekri Ershad, Shiraz International University, Iran
Mr. Md. Safiqul Islam, Daffodil International University, Bangladesh
Mr. Mahmudul Hasan, Daffodil International University, Bangladesh
Prof. Mandakini Tayade, UIT, RGTU, Bhopal, India
Ms. Sarla More, UIT, RGTU, Bhopal, India
Mr. Tushar Hrishikesh Jaware, R.C. Patel Institute of Technology, Shirpur, India
Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore, India
Mr. Fahimuddin Shaik, Annamacharya Institute of Technology & Sciences, India
Dr. M. N. Giri Prasad, JNTUCE, Pulivendula, A.P., India
Assist. Prof. Chintan M Bhatt, Charotar University of Science And Technology, India
Prof. Sahista Machchhar, Marwadi Education Foundation's Group of institutions, India
Assist. Prof. Navnish Goel, S. D. College Of Engineering & Technology, India
Mr. Khaja Kamaluddin, Sirt University, Sirt, Libya
Mr. Mohammad Zaidul Karim, Daffodil International, Bangladesh
Mr. M. Vijayakumar, KSR College of Engineering, Tiruchengode, India
Mr. S. A. Ahsan Rajon, Khulna University, Bangladesh
Dr. Muhammad Mohsin Nazir, LCW University Lahore, Pakistan
Mr. Mohammad Asadul Hoque, University of Alabama, USA
Mr. P.V.Sarathchand, Indur Institute of Engineering and Technology, India
Mr. Durgesh Samadhiya, Chung Hua University, Taiwan
Dr Venu Kuthadi, University of Johannesburg, Johannesburg, RSA
Dr. (Er) Jasvir Singh, Guru Nanak Dev University, Amritsar, Punjab, India
Mr. Jasmin Cosic, Min. of the Interior of Una-sana canton, B&H, Bosnia and Herzegovina
Dr S. Rajalakshmi, Botho College, South Africa
Dr. Mohamed Sarrab, De Montfort University, UK
Mr. Basappa B. Kodada, Canara Engineering College, India
Assist. Prof. K. Ramana, Annamacharya Institute of Technology and Sciences, India
Dr. Ashu Gupta, Apeejay Institute of Management, Jalandhar, India
Assist. Prof. Shaik Rasool, Shadan College of Engineering & Technology, India
Assist. Prof. K. Suresh, Annamacharya Institute of Tech & Sci. Rajampet, AP, India
Dr . G. Singaravel, K.S.R. College of Engineering, India
Dr B. G. Geetha, K.S.R. College of Engineering, India
Assist. Prof. Kavita Choudhary, ITM University, Gurgaon
Dr. Mehrdad Jalali, Azad University, Mashhad, Iran
Megha Goel, Shamli Institute of Engineering and Technology, Shamli, India
Mr. Chi-Hua Chen, Institute of Information Management, National Chiao-Tung University, Taiwan (R.O.C.)
Assoc. Prof. A. Rajendran, RVS College of Engineering and Technology, India
Assist. Prof. S. Jaganathan, RVS College of Engineering and Technology, India
Assoc. Prof. (Dr.) A S N Chakravarthy, JNTUK University College of Engineering Vizianagaram (State University)
Assist. Prof. Deepshikha Patel, Technocrat Institute of Technology, India
Assist. Prof. Maram Balajee, GMRIT, India
Assist. Prof. Monika Bhatnagar, TIT, India
Prof. Gaurang Panchal, Charotar University of Science & Technology, India
Prof. Anand K. Tripathi, Computer Society of India
Prof. Jyoti Chaudhary, High Performance Computing Research Lab, India
Assist. Prof. Supriya Raheja, ITM University, India

Dr. Pankaj Gupta, Microsoft Corporation, U.S.A.
Assist. Prof. Panchamukesh Chandaka, Hyderabad Institute of Tech. & Management, India
Prof. Mohan H.S, SJB Institute Of Technology, India
Mr. Hossein Malekinezhad, Islamic Azad University, Iran
Mr. Zatin Gupta, Universti Malaysia, Malaysia
Assist. Prof. Amit Chauhan, Phonics Group of Institutions, India
Assist. Prof. Ajal A. J., METS School Of Engineering, India
Mrs. Omowunmi Omobola Adeyemo, University of Ibadan, Nigeria
Dr. Bharat Bhushan Agarwal, I.F.T.M. University, India
Md. Nazrul Islam, University of Western Ontario, Canada
Tushar Kanti, L.N.C.T, Bhopal, India
Er. Aumreesh Kumar Saxena, SIRTs College Bhopal, India
Mr. Mohammad Monirul Islam, Daffodil International University, Bangladesh
Dr. Kashif Nisar, University Utara Malaysia, Malaysia
Dr. Wei Zheng, Rutgers Univ/ A10 Networks, USA
Associate Prof. Rituraj Jain, Vyas Institute of Engg & Tech, Jodhpur – Rajasthan
Assist. Prof. Apoorvi Sood, I.T.M. University, India
Dr. Kayhan Zrar Ghafoor, University Technology Malaysia, Malaysia
Mr. Swapnil Sonar, Truba Institute College of Engineering & Technology, Indore, India
Ms. Yogita Gigras, I.T.M. University, India
Associate Prof. Neelima Sadineni, Pydha Engineering College, India Pydha Engineering College
Assist. Prof. K. Deepika Rani, HITAM, Hyderabad
Ms. Shikha Maheshwari, Jaipur Engineering College & Research Centre, India
Prof. Dr V S Giridhar Akula, Avanthi's Scientific Tech. & Research Academy, Hyderabad
Prof. Dr.S.Saravanan, Muthayammal Engineering College, India
Mr. Mehdi Golsorkhatabar Amiri, Islamic Azad University, Iran
Prof. Amit Sadanand Savyanavar, MITCOE, Pune, India
Assist. Prof. P.Oliver Jayaprakash, Anna University, Chennai
Assist. Prof. Ms. Sujata, ITM University, Gurgaon, India
Dr. Asoke Nath, St. Xavier's College, India
Mr. Masoud Rafiqhi, Islamic Azad University, Iran
Assist. Prof. RamBabu Pemula, NIMRA College of Engineering & Technology, India
Assist. Prof. Ms Rita Chhikara, ITM University, Gurgaon, India
Mr. Sandeep Maan, Government Post Graduate College, India
Prof. Dr. S. Muralidharan, Mepco Schlenk Engineering College, India
Associate Prof. T.V.Sai Krishna, QIS College of Engineering and Technology, India
Mr. R. Balu, Bharathiar University, Coimbatore, India
Assist. Prof. Shekhar. R, Dr.SM College of Engineering, India
Prof. P. Senthilkumar, Vivekanandha Institue of Engineering and Technology for Woman, India
Mr. M. Kamarajan, PSNA College of Engineering & Technology, India
Dr. Angajala Srinivasa Rao, Jawaharlal Nehru Technical University, India
Assist. Prof. C. Venkatesh, A.I.T.S, Rajampet, India
Mr. Afshin Rezakhani Roozbahani, Ayatollah Boroujerdi University, Iran
Mr. Laxmi chand, SCTL, Noida, India
Dr. Dr. Abdul Hannan, Vivekanand College, Aurangabad
Prof. Mahesh Panchal, KITRC, Gujarat
Dr. A. Subramani, K.S.R. College of Engineering, Tiruchengode

Assist. Prof. Prakash M, Rajalakshmi Engineering College, Chennai, India
Assist. Prof. Akhilesh K Sharma, Sir Padampat Singhanian University, India
Ms. Varsha Sahni, Guru Nanak Dev Engineering College, Ludhiana, India
Associate Prof. Trilochan Rout, NM Institute of Engineering and Technology, India
Mr. Srikanta Kumar Mohapatra, NMIET, Orissa, India
Mr. Waqas Haider Bangyal, Iqra University Islamabad, Pakistan
Dr. S. Vijayaragavan, Christ College of Engineering and Technology, Pondicherry, India
Prof. Elboukhari Mohamed, University Mohammed First, Oujda, Morocco
Dr. Muhammad Asif Khan, King Faisal University, Saudi Arabia
Dr. Nagy Ramadan Darwish Omran, Cairo University, Egypt.
Assistant Prof. Anand Nayyar, KCL Institute of Management and Technology, India
Mr. G. Premsankar, Ericsson, India
Assist. Prof. T. Hemalatha, VELS University, India
Prof. Tejaswini Apte, University of Pune, India
Dr. Edmund Ng Giap Weng, Universiti Malaysia Sarawak, Malaysia
Mr. Mahdi Nouri, Iran University of Science and Technology, Iran
Associate Prof. S. Asif Hussain, Annamacharya Institute of technology & Sciences, India
Mrs. Kavita Pabreja, Maharaja Surajmal Institute (an affiliate of GGSIP University), India
Mr. Vorugunti Chandra Sekhar, DA-IICT, India
Mr. Muhammad Najmi Ahmad Zabidi, Universiti Teknologi Malaysia, Malaysia
Dr. Aderemi A. Atayero, Covenant University, Nigeria
Assist. Prof. Osama Sohaib, Balochistan University of Information Technology, Pakistan
Assist. Prof. K. Suresh, Annamacharya Institute of Technology and Sciences, India
Mr. Hassen Mohammed Abdullaah Alsafi, International Islamic University Malaysia (IIUM) Malaysia
Mr. Robail Yasrab, Virtual University of Pakistan, Pakistan
Mr. R. Balu, Bharathiar University, Coimbatore, India
Prof. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar
Assoc. Prof. Vivek S Deshpande, MIT College of Engineering, India
Prof. K. Saravanan, Anna university Coimbatore, India
Dr. Ravendra Singh, MJP Rohilkhand University, Bareilly, India
Mr. V. Mathivanan, IBRA College of Technology, Sultanate of OMAN
Assoc. Prof. S. Asif Hussain, AITS, India
Assist. Prof. C. Venkatesh, AITS, India
Mr. Sami Ulhaq, SZABIST Islamabad, Pakistan
Dr. B. Justus Rabi, Institute of Science & Technology, India
Mr. Anuj Kumar Yadav, Dehradun Institute of technology, India
Mr. Alejandro Mosquera, University of Alicante, Spain
Assist. Prof. Arjun Singh, Sir Padampat Singhanian University (SPSU), Udaipur, India
Dr. Smriti Agrawal, JB Institute of Engineering and Technology, Hyderabad
Assist. Prof. Swathi Sambangi, Visakha Institute of Engineering and Technology, India
Ms. Prabhjot Kaur, Guru Gobind Singh Indraprastha University, India
Mrs. Samaher AL-Hothali, Yanbu University College, Saudi Arabia
Prof. Rajneeshkaur Bedi, MIT College of Engineering, Pune, India
Mr. Hassen Mohammed Abdullaah Alsafi, International Islamic University Malaysia (IIUM)
Dr. Wei Zhang, Amazon.com, Seattle, WA, USA
Mr. B. Santhosh Kumar, C S I College of Engineering, Tamil Nadu
Dr. K. Reji Kumar, N S S College, Pandalam, India

Assoc. Prof. K. Seshadri Sastry, EILM University, India
Mr. Kai Pan, UNC Charlotte, USA
Mr. Ruikar Sachin, SGGSIET, India
Prof. (Dr.) Vinodani Katiyar, Sri Ramswaroop Memorial University, India
Assoc. Prof., M. Giri, Sreenivasa Institute of Technology and Management Studies, India
Assoc. Prof. Labib Francis Gergis, Misr Academy for Engineering and Technology (MET), Egypt
Assist. Prof. Amanpreet Kaur, ITM University, India
Assist. Prof. Anand Singh Rajawat, Shri Vaishnav Institute of Technology & Science, Indore
Mrs. Hadeel Saleh Haj Aliwi, Universiti Sains Malaysia (USM), Malaysia
Dr. Abhay Bansal, Amity University, India
Dr. Mohammad A. Mezher, Fahad Bin Sultan University, KSA
Assist. Prof. Nidhi Arora, M.C.A. Institute, India
Prof. Dr. P. Suresh, Karpagam College of Engineering, Coimbatore, India
Dr. Kannan Balasubramanian, Mepco Schlenk Engineering College, India
Dr. S. Sankara Gomathi, Panimalar Engineering college, India
Prof. Anil kumar Suthar, Gujarat Technological University, L.C. Institute of Technology, India
Assist. Prof. R. Hubert Rajan, NOORUL ISLAM UNIVERSITY, India
Assist. Prof. Dr. Jyoti Mahajan, College of Engineering & Technology
Assist. Prof. Homam Reda El-Taj, College of Network Engineering, Saudi Arabia & Malaysia
Mr. Bijan Paul, Shahjalal University of Science & Technology, Bangladesh
Assoc. Prof. Dr. Ch V Phani Krishna, KL University, India
Dr. Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies & Research, India
Dr. Lamir LAOUAMER, Al Qassim University, Dept. Info. Systems & European University of Brittany, Dept. Computer Science, UBO, Brest, France
Prof. Ashish Babanrao Sasankar, G.H.Raisoni Institute Of Information Technology, India
Prof. Pawan Kumar Goel, Shamli Institute of Engineering and Technology, India
Mr. Ram Kumar Singh, S.V Subharti University, India
Assistant Prof. Sunish Kumar O S, Amalijothei College of Engineering, India
Dr Sanjay Bhargava, Banasthali University, India
Mr. Pankaj S. Kulkarni, AVEW's Shatabdi Institute of Technology, India
Mr. Roohollah Etemadi, Islamic Azad University, Iran
Mr. Oloruntoyin Sefiu Taiwo, Emmanuel Alayande College Of Education, Nigeria
Mr. Sumit Goyal, National Dairy Research Institute, India
Mr Jaswinder Singh Dilawari, Geeta Engineering College, India
Prof. Raghuraj Singh, Harcourt Butler Technological Institute, Kanpur
Dr. S.K. Mahendran, Anna University, Chennai, India
Dr. Amit Wason, Hindustan Institute of Technology & Management, Punjab
Dr. Ashu Gupta, Apeejay Institute of Management, India
Assist. Prof. D. Asir Antony Gnana Singh, M.I.E.T Engineering College, India
Mrs Mina Farmanbar, Eastern Mediterranean University, Famagusta, North Cyprus
Mr. Maram Balajee, GMR Institute of Technology, India
Mr. Moiz S. Ansari, Isra University, Hyderabad, Pakistan
Mr. Adebayo, Olawale Surajudeen, Federal University of Technology Minna, Nigeria
Mr. Jasvir Singh, University College Of Engg., India
Mr. Vivek Tiwari, MANIT, Bhopal, India
Assoc. Prof. R. Navaneethakrishnan, Bharathiyar College of Engineering and Technology, India
Mr. Somdip Dey, St. Xavier's College, Kolkata, India

Mr. Souleymane Balla-Arabé, Xi'an University of Electronic Science and Technology, China
Mr. Mahabub Alam, Rajshahi University of Engineering and Technology, Bangladesh
Mr. Sathyapraksh P., S.K.P Engineering College, India
Dr. N. Karthikeyan, SNS College of Engineering, Anna University, India
Dr. Binod Kumar, JSPM's, Jayawant Technical Campus, Pune, India
Assoc. Prof. Dinesh Goyal, Suresh Gyan Vihar University, India
Mr. Md. Abdul Ahad, K L University, India
Mr. Vikas Bajpai, The LNM IIT, India
Dr. Manish Kumar Anand, Salesforce (R & D Analytics), San Francisco, USA
Assist. Prof. Dheeraj Murari, Kumaon Engineering College, India
Assoc. Prof. Dr. A. Muthukumaravel, VELS University, Chennai
Mr. A. Siles Balasingh, St. Joseph University in Tanzania, Tanzania
Mr. Ravindra Daga Badgujar, R C Patel Institute of Technology, India
Dr. Preeti Khanna, SVKM's NMIMS, School of Business Management, India
Mr. Kumar Dayanand, Cambridge Institute of Technology, India
Dr. Syed Asif Ali, SMI University Karachi, Pakistan
Prof. Pallvi Pandit, Himachal Pradesh University, India
Mr. Ricardo Verschueren, University of Gloucestershire, UK
Assist. Prof. Mamta Juneja, University Institute of Engineering and Technology, Panjab University, India
Assoc. Prof. P. Surendra Varma, NRI Institute of Technology, JNTU Kakinada, India
Assist. Prof. Gaurav Shrivastava, RGPV / SVITS Indore, India
Dr. S. Sumathi, Anna University, India
Assist. Prof. Ankita M. Kapadia, Charotar University of Science and Technology, India
Mr. Deepak Kumar, Indian Institute of Technology (BHU), India
Dr. Dr. Rajan Gupta, GGSIP University, New Delhi, India
Assist. Prof. M. Anand Kumar, Karpagam University, Coimbatore, India
Mr. Mr Arshad Mansoor, Pakistan Aeronautical Complex
Mr. Kapil Kumar Gupta, Ansal Institute of Technology and Management, India
Dr. Neeraj Tomer, SINE International Institute of Technology, Jaipur, India
Assist. Prof. Trunal J. Patel, C.G.Patel Institute of Technology, Uka Tarsadia University, Bardoli, Surat
Mr. Sivakumar, Codework solutions, India
Mr. Mohammad Sadegh Mirzaei, PGNR Company, Iran
Dr. Gerard G. Dumancas, Oklahoma Medical Research Foundation, USA
Mr. Varadala Sridhar, Varadhman College Engineering College, Affiliated To JNTU, Hyderabad
Assist. Prof. Manoj Dhawan, SVITS, Indore
Assoc. Prof. Chitreshh Banerjee, Suresh Gyan Vihar University, Jaipur, India
Dr. S. Santhi, SCSVMV University, India
Mr. Davood Mohammadi Souran, Ministry of Energy of Iran, Iran
Mr. Shamim Ahmed, Bangladesh University of Business and Technology, Bangladesh
Mr. Sandeep Reddivari, Mississippi State University, USA
Assoc. Prof. Ousmane Thiare, Gaston Berger University, Senegal
Dr. Hazra Imran, Athabasca University, Canada
Dr. Setu Kumar Chaturvedi, Technocrats Institute of Technology, Bhopal, India
Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology, India
Ms. Jaspreet Kaur, Distance Education LPU, India
Dr. D. Nagarajan, Salalah College of Technology, Sultanate of Oman
Dr. K.V.N.R.Sai Krishna, S.V.R.M. College, India

Mr. Himanshu Pareek, Center for Development of Advanced Computing (CDAC), India
Mr. Khaldi Amine, Badji Mokhtar University, Algeria
Mr. Mohammad Sadegh Mirzaei, Scientific Applied University, Iran
Assist. Prof. Khyati Chaudhary, Ram-eesh Institute of Engg. & Technology, India
Mr. Sanjay Agal, Pacific College of Engineering Udaipur, India
Mr. Abdul Mateen Ansari, King Khalid University, Saudi Arabia
Dr. H.S. Behera, Veer Surendra Sai University of Technology (VSSUT), India
Dr. Shrikant Tiwari, Shri Shankaracharya Group of Institutions (SSGI), India
Prof. Ganesh B. Regulwar, Shri Shankarprasad Agnihotri College of Engg, India
Prof. Pinnamaneni Bhanu Prasad, Matrix vision GmbH, Germany
Dr. Shrikant Tiwari, Shri Shankaracharya Technical Campus (SSTC), India
Dr. Siddesh G.K., : Dayananada Sagar College of Engineering, Bangalore, India
Dr. Nadir Bouchama, CERIST Research Center, Algeria
Dr. R. Sathishkumar, Sri Venkateswara College of Engineering, India
Assistant Prof (Dr.) Mohamed Moussaoui, Abdelmalek Essaadi University, Morocco
Dr. S. Malathi, Panimalar Engineering College, Chennai, India
Dr. V. Subedha, Panimalar Institute of Technology, Chennai, India
Dr. Prashant Panse, Swami Vivekanand College of Engineering, Indore, India
Dr. Hamza Aldabbas, Al-Balqa'a Applied University, Jordan
Dr. G. Rasitha Banu, Vel's University, Chennai
Dr. V. D. Ambeth Kumar, Panimalar Engineering College, Chennai
Prof. Anuranjan Misra, Bhagwant Institute of Technology, Ghaziabad, India
Ms. U. Sinthuja, PSG college of arts & science, India
Dr. Ehsan Saradar Torshizi, Urmia University, Iran
Dr. Shamneesh Sharma, APG Shimla University, Shimla (H.P.), India
Assistant Prof. A. S. Syed Navaz, Muthayammal College of Arts & Science, India
Assistant Prof. Ranjit Panigrahi, Sikkim Manipal Institute of Technology, Majitar, Sikkim
Dr. Khaled Eskaf, Arab Academy for Science ,Technology & Maritime Transportation, Egypt
Dr. Nishant Gupta, University of Jammu, India
Assistant Prof. Nagarajan Sankaran, Annamalai University, Chidambaram, Tamilnadu, India
Assistant Prof. Tribikram Pradhan, Manipal Institute of Technology, India
Dr. Nasser Lotfi, Eastern Mediterranean University, Northern Cyprus
Dr. R. Manavalan, K S Rangasamy college of Arts and Science, Tamilnadu, India
Assistant Prof. P. Krishna Sankar, K S Rangasamy college of Arts and Science, Tamilnadu, India
Dr. Rahul Malik, Cisco Systems, USA
Dr. S. C. Lingareddy, ALPHA College of Engineering, India
Assistant Prof. Mohammed Shuaib, Interat University, Lucknow, India
Dr. Sachin Yele, Sanghvi Institute of Management & Science, India
Dr. T. Thambidurai, Sun Univercell, Singapore
Prof. Anandkumar Telang, BKIT, India
Assistant Prof. R. Poorvadevi, SCSVMV University, India
Dr Uttam Mande, Gitam University, India
Dr. Poornima Girish Naik, Shahu Institute of Business Education and Research (SIBER), India
Prof. Md. Abu Kausar, Jaipur National University, Jaipur, India
Dr. Mohammed Zuber, AISECT University, India
Prof. Kalum Priyanath Udagepola, King Abdulaziz University, Saudi Arabia
Dr. K. R. Ananth, Velalar College of Engineering and Technology, India

Assistant Prof. Sanjay Sharma, Roorkee Engineering & Management Institute Shamli (U.P), India
Assistant Prof. Panem Charan Arur, Priyadarshini Institute of Technology, India
Dr. Ashwak Mahmood muhsen alabaichi, Karbala University / College of Science, Iraq
Dr. Urmila Shrawankar, G H Raison College of Engineering, Nagpur (MS), India
Dr. Krishan Kumar Paliwal, Panipat Institute of Engineering & Technology, India
Dr. Mukesh Negi, Tech Mahindra, India
Dr. Anuj Kumar Singh, Amity University Gurgaon, India
Dr. Babar Shah, Gyeongsang National University, South Korea
Assistant Prof. Jayprakash Upadhyay, SRI-TECH Jabalpur, India
Assistant Prof. Varadala Sridhar, Vidya Jyothi Institute of Technology, India
Assistant Prof. Parameshachari B D, KSIT, Bangalore, India
Assistant Prof. Ankit Garg, Amity University, Haryana, India
Assistant Prof. Rajashe Karappa, SDMCET, Karnataka, India
Assistant Prof. Varun Jasuja, GNIT, India
Assistant Prof. Sonal Honale, Abha Gaikwad Patil College of Engineering Nagpur, India
Dr. Pooja Choudhary, CT Group of Institutions, NIT Jalandhar, India
Dr. Faouzi Hidoussi, UHL Batna, Algeria
Dr. Naseer Ali Hussein, Wasit University, Iraq
Assistant Prof. Vinod Kumar Shukla, Amity University, Dubai
Dr. Ahmed Farouk Metwaly, K L University
Mr. Mohammed Noaman Murad, Cihan University, Iraq
Dr. Suxing Liu, Arkansas State University, USA
Dr. M. Gomathi, Velalar College of Engineering and Technology, India
Assistant Prof. Sumardiono, College PGRI Blitar, Indonesia
Dr. Latika Kharb, Jagan Institute of Management Studies (JIMS), Delhi, India
Associate Prof. S. Raja, Pauls College of Engineering and Technology, Tamilnadu, India
Assistant Prof. Seyed Reza Pakize, Shahid Sani High School, Iran
Dr. Thiyagu Nagaraj, University-INO, India
Assistant Prof. Noreen Sarai, Harare Institute of Technology, Zimbabwe
Assistant Prof. Gajanand Sharma, Suresh Gyan Vihar University Jaipur, Rajasthan, India
Assistant Prof. Mapari Vikas Prakash, Siddhant COE, Sudumbare, Pune, India
Dr. Devesh Katiyar, Shri Ramswaroop Memorial University, India
Dr. Shenshen Liang, University of California, Santa Cruz, US
Assistant Prof. Mohammad Abu Omar, Limkokwing University of Creative Technology- Malaysia
Mr. Snehasis Banerjee, Tata Consultancy Services, India
Assistant Prof. Kibona Lusekelo, Ruaha Catholic University (RUCU), Tanzania
Assistant Prof. Adib Kabir Chowdhury, University College Technology Sarawak, Malaysia
Dr. Ying Yang, Computer Science Department, Yale University, USA
Dr. Vinay Shukla, Institute Of Technology & Management, India
Dr. Liviu Octavian Maftciu-Scai, West University of Timisoara, Romania
Assistant Prof. Rana Khudhair Abbas Ahmed, Al-Rafidain University College, Iraq
Assistant Prof. Nitin A. Naik, S.R.T.M. University, India
Dr. Timothy Powers, University of Hertfordshire, UK
Dr. S. Prasath, Bharathiar University, Erode, India
Dr. Ritu Shrivastava, SIRTIS Bhopal, India
Prof. Rohit Shrivastava, Mittal Institute of Technology, Bhopal, India
Dr. Gianina Mihai, Dunarea de Jos" University of Galati, Romania

Assistant Prof. Ms. T. Kalai Selvi, Erode Sengunthar Engineering College, India
Assistant Prof. Ms. C. Kavitha, Erode Sengunthar Engineering College, India
Assistant Prof. K. Sinivasamoorthi, Erode Sengunthar Engineering College, India
Assistant Prof. Mallikarjun C Sarsamba Bheemna Khandre Institute Technology, Bhalki, India
Assistant Prof. Vishwanath Chikaraddi, Veermata Jijabai technological Institute (Central Technological Institute), India
Assistant Prof. Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, India
Assistant Prof. Mohammed Noaman Murad, Cihan University, Iraq
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco
Dr. Parul Verma, Amity University, India
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco
Assistant Prof. Madhavi Dhingra, Amity University, Madhya Pradesh, India
Assistant Prof.. G. Selvavinayagam, SNS College of Technology, Coimbatore, India
Assistant Prof. Madhavi Dhingra, Amity University, MP, India
Professor Kartheesan Log, Anna University, Chennai
Professor Vasudeva Acharya, Shri Madhwa vadiraja Institute of Technology, India
Dr. Asif Iqbal Hajamydeen, Management & Science University, Malaysia
Assistant Prof., Mahendra Singh Meena, Amity University Haryana
Assistant Professor Manjeet Kaur, Amity University Haryana
Dr. Mohamed Abd El-Basset Matwalli, Zagazig University, Egypt
Dr. Ramani Kannan, Universiti Teknologi PETRONAS, Malaysia
Assistant Prof. S. Jagadeesan Subramaniam, Anna University, India
Assistant Prof. Dharmendra Choudhary, Tripura University, India
Assistant Prof. Deepika Vodnala, SR Engineering College, India
Dr. Kai Cong, Intel Corporation & Computer Science Department, Portland State University, USA
Dr. Kailas R Patil, Vishwakarma Institute of Information Technology (VIIT), India
Dr. Omar A. Alzubi, Faculty of IT / Al-Balqa Applied University, Jordan
Assistant Prof. Kareemullah Shaik, Nimra Institute of Science and Technology, India
Assistant Prof. Chirag Modi, NIT Goa
Dr. R. Ramkumar, Nandha Arts And Science College, India
Dr. Priyadarshini Vydhialingam, Harathiar University, India
Dr. P. S. Jagadeesh Kumar, DBIT, Bangalore, Karnataka
Dr. Vikas Thada, AMITY University, Pachgaon
Dr. T. A. Ashok Kumar, Institute of Management, Christ University, Bangalore
Dr. Shaheera Rashwan, Informatics Research Institute
Dr. S. Preetha Gunasekar, Bharathiyar University, India
Asst Professor Sameer Dev Sharma, Uttaranchal University, Dehradun
Dr. Zhihan Iv, Chinese Academy of Science, China
Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, Amritsar
Dr. Umar Ruhi, University of Ottawa, Canada
Dr. Jasmin Cosic, University of Bihac, Bosnia and Herzegovina
Dr. Homam Reda El-Taj, University of Tabuk, Kingdom of Saudi Arabia
Dr. Mostafa Ghobaei Arani, Islamic Azad University, Iran
Dr. Ayyasamy Ayyanar, Annamalai University, India
Dr. Selvakumar Manickam, Universiti Sains Malaysia, Malaysia
Dr. Murali Krishna Namana, GITAM University, India
Dr. Smriti Agrawal, Chaitanya Bharathi Institute of Technology, Hyderabad, India
Professor Vimalathithan Rathinasabapathy, Karpagam College Of Engineering, India

Dr. Sushil Chandra Dimri, Graphic Era University, India
Dr. Dinh-Sinh Mai, Le Quy Don Technical University, Vietnam
Dr. S. Rama Sree, Aditya Engg. College, India
Dr. Ehab T. Alnfrawy, Sadat Academy, Egypt
Dr. Patrick D. Cerna, Haramaya University, Ethiopia
Dr. Vishal Jain, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), India
Associate Prof. Dr. Jiliang Zhang, North Eastern University, China
Dr. Sharefa Murad, Middle East University, Jordan
Dr. Ajeet Singh Poonia, Govt. College of Engineering & technology, Rajasthan, India
Dr. Vahid Esmaeaelzadeh, University of Science and Technology, Iran
Dr. Jacek M. Czerniak, Casimir the Great University in Bydgoszcz, Institute of Technology, Poland
Associate Prof. Anisur Rehman Nasir, Jamia Millia Islamia University
Assistant Prof. Imran Ahmad, COMSATS Institute of Information Technology, Pakistan
Professor Ghulam Qasim, Preston University, Islamabad, Pakistan
Dr. Parameshachari B D, GSSS Institute of Engineering and Technology for Women
Dr. Wencan Luo, University of Pittsburgh, US
Dr. Musa PEKER, Faculty of Technology, Mugla Sitki Kocman University, Turkey
Dr. Gunasekaran Shanmugam, Anna University, India
Dr. Binh P. Nguyen, National University of Singapore, Singapore
Dr. Rajkumar Jain, Indian Institute of Technology Indore, India
Dr. Imtiaz Ali Halepoto, QUEST Nawabshah, Pakistan
Dr. Shaligram Prajapat, Devi Ahilya University Indore India
Dr. Sunita Singhal, Birla Institute of Technology and Science, Pilani, India
Dr. Ijaz Ali Shoukat, King Saud University, Saudi Arabia
Dr. Anuj Gupta, IKG Punjab Technical University, India
Dr. Sonali Saini, IES-IPS Academy, India
Dr. Krishan Kumar, MotiLal Nehru National Institute of Technology, Allahabad, India
Dr. Z. Faizal Khan, College of Engineering, Shaqra University, Kingdom of Saudi Arabia
Prof. M. Padmavathamma, S.V. University Tirupati, India
Prof. A. Velayudham, Cape Institute of Technology, India
Prof. Seifeidne Kadry, American University of the Middle East
Dr. J. Durga Prasad Rao, Pt. Ravishankar Shukla University, Raipur
Assistant Prof. Najam Hasan, Dhofar University
Dr. G. Suseendran, Vels University, Pallavaram, Chennai
Prof. Ankit Faldu, Gujarat Technological University- Atmiya Institute of Technology and Science
Dr. Ali Habiboghli, Islamic Azad University
Dr. Deepak Dembla, JECRC University, Jaipur, India
Dr. Pankaj Rajan, Walmart Labs, USA
Assistant Prof. Radoslava Kraveva, South-West University "Neofit Rilski", Bulgaria
Assistant Prof. Medhavi Shriwas, Shri vaishnav institute of Technology, India
Associate Prof. Sedat Akleylek, Ondokuz Mayıs University, Turkey
Dr. U.V. Arivazhagu, Kingston Engineering College Affiliated To Anna University, India
Dr. Touseef Ali, University of Engineering and Technology, Taxila, Pakistan
Assistant Prof. Naren Jeeva, SASTRA University, India
Dr. Riccardo Colella, University of Salento, Italy
Dr. Enache Maria Cristina, University of Galati, Romania
Dr. Senthil P, Kurinji College of Arts & Science, India

Dr. Hasan Ashrafi-rizi, Isfahan University of Medical Sciences, Isfahan, Iran
Dr. Mazhar Malik, Institute of Southern Punjab, Pakistan
Dr. Yajie Miao, Carnegie Mellon University, USA
Dr. Kamran Shaukat, University of the Punjab, Pakistan
Dr. Sasikaladevi N., SASTRA University, India
Dr. Ali Asghar Rahmani Hosseinabadi, Islamic Azad University Ayatollah Amoli Branch, Amol, Iran
Dr. Velin Krlev, South-West University "Neofit Rilski", Blagoevgrad, Bulgaria
Dr. Marius Iulian Mihailescu, LUMINA - The University of South-East Europe
Dr. Sriramula Nagaprasad, S.R.R.Govt.Arts & Science College, Karimnagar, India
Prof (Dr.) Namrata Dhanda, Dr. APJ Abdul Kalam Technical University, Lucknow, India
Dr. Javed Ahmed Mahar, Shah Abdul Latif University, Khairpur Mir's, Pakistan
Dr. B. Narendra Kumar Rao, Sree Vidyanikethan Engineering College, India
Dr. Shahzad Anwar, University of Engineering & Technology Peshawar, Pakistan
Dr. Basit Shahzad, King Saud University, Riyadh - Saudi Arabia
Dr. Nilamadhab Mishra, Chang Gung University
Dr. Sachin Kumar, Indian Institute of Technology Roorkee
Dr. Santosh Nanda, Biju-Pattnaik University of Technology
Dr. Sherzod Turaev, International Islamic University Malaysia
Dr. Yilun Shang, Tongji University, Department of Mathematics, Shanghai, China
Dr. Nuzhat Shaikh, Modern Education society's College of Engineering, Pune, India
Dr. Parul Verma, Amity University, Lucknow campus, India
Dr. Rachid Alaoui, Agadir Ibn Zohr University, Agadir, Morocco
Dr. Dharmendra Patel, Charotar University of Science and Technology, India
Dr. Dong Zhang, University of Central Florida, USA
Dr. Kennedy Chinedu Okafor, Federal University of Technology Owerri, Nigeria
Prof. C Ram Kumar, Dr NGP Institute of Technology, India
Dr. Sandeep Gupta, GGS IP University, New Delhi, India
Dr. Shahanawaj Ahamad, University of Ha'il, Ha'il City, Ministry of Higher Education, Kingdom of Saudi Arabia
Dr. Najeed Ahmed Khan, NED University of Engineering & Technology, India
Dr. Sajid Ullah Khan, Universiti Malaysia Sarawak, Malaysia
Dr. Muhammad Asif, National Textile University Faisalabad, Pakistan
Dr. Yu BI, University of Central Florida, Orlando, FL, USA
Dr. Brijendra Kumar Joshi, Research Center, Military College of Telecommunication Engineering, India
Prof. Dr. Nak Eun Cho, Pukyong National University, Korea
Prof. Wasim Ul-Haq, Mathematics Department Faculty of Science, Majmaah University, Saudi Arabia

CALL FOR PAPERS
International Journal of Computer Science and Information Security

IJCSIS 2017-2018

ISSN: 1947-5500

<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .



© IJCSIS PUBLICATION 2017

ISSN 1947 5500

<http://sites.google.com/site/ijcsis/>